DOCUMENT RESUME

ED 264 542 CS 008 245

AUTHOR	Niles, Jerome A., Ed.; Harris, Larry A., Ed.
TITLE	New Inquiries in Reading Research and Instruction. Thirty-First Yearbook of the National Reading Conference.
INSTITUTION	National Reading Conference, Inc.
PUB DATE	82
NOTE	324p.; Proceedings of the Annual Meeting of the National Reading Conference (31st, Dallas, TX, December 2-5, 1981).
PUB TYPE	Books (010) -- Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	Beginning Reading; Classroom Environment; Primary Education; Reading Comprehension; Reading Diagnosis; *Reading Instruction; *Reading Research; Teacher Effectiveness; Teaching Methods; Vocabulary Development
IDENTIFIERS	Story Structure

ABSTRACT

Reflecting an intensified concern for good educational research on instruction, the papers in this collection demonstrate the growing awareness of and sensitivity to the classroom environment by reading researchers. The 51 articles, selected from those presented at the 1981 meeting of the National Reading Conference, are divided into the following categories: (1) research on reading instruction; (2) reading comprehension: learning and instruction; (3) comprehension: vocabulary; (4) beginning reading; (5) research and measurement concerns; (6) teacher effectiveness; (7) review of research on technology and reading instruction; (8) story structure; and (9) writing. The collection concludes with the complete list of conference papers from the 1981 meeting. (HTH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

[✓ This document has been reproduced as received from the person or organization originating it.]

Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED264542

NEW INQUIRIES IN READING RESEARCH AND INSTRUCTION

Thirty-first Yearbook
of
The National Reading Conference

JEROME A. NILES, Editor
and
LARRY A. HARRIS, Associate Editor
Virginia Polytechnic Institute and State University

Published by
The National Reading Conference, Inc.

1982

CS008245

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
Peter O. Allen

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Copyright, 1982
THE NATIONAL READING CONFERENCE, INC.
All rights reserved. This book or parts thereof
may not be reproduced in any form without
permission of the publishers.

Distributed by
THE NATIONAL READING CONFERENCE
1070 Sibley Tower
Rochester, N.Y. 14604

The ideas expressed in the papers published
in the yearbook are those of the authors and do
not necessarily represent the viewpoints
of the National Reading Conference.

Information concerning membership in The National Reading Conference and subscriptions to *Journal of Reading Behavior* and the *NRC Yearbook* may be obtained by writing to:

National Reading Conference
1070 Sibley Tower
Rochester, N.Y. 14604

OFFICERS AND BOARD OF DIRECTORS
OF THE
NATIONAL READING CONFERENCE, INC.

FRANK GREENE, *President*
McGill University

IRENE ATHEY, *Vice-President*
Rutgers University

M. TRIKA SMITH-BURKE, *Treasurer*
New York University

NANCY MARSHALL, *Secretary*
Florida International University

HARRY SINGER, *Past President*
University of California at Riverside

JOHN T. GUTHRIE, *Board Member*
International Reading Assn.

JAMES V. HOFFMAN, *Board Member*
University of Texas-Austin

P. DAVID PEARSON, *Board Member*
University of Illinois

JOANNA P. WILLIAMS, *Board Member*
Columbia University

MICHAEL McKENNA, *Convention Manager*
Wichita State University

PRISCILLA DRUM, *Publications Chairperson*
University of California

PETER O. ALLEN, *Executive Director*

EDITORIAL ADVISORY BOARD

NRC Yearbook

Patricia L. Anders
University of Arizona

Margaret Atwell
California State College

Mark W. Aulls
McGill University

Hilda Borko
VPI & SU

James W. Cunningham
University of North Carolina

Patricia Cummingham
Wake Forest University

Mariam Jean Dreher
University of Maryland

Priscilla A. Drum
University of California

Gerald D. Duffy
Michigan State University

Josephine Spivack Goldsmith
Rutgers University

Jane Hansen
University of New Hampshire

Jerome C. Harste
Indiana University

James V. Hoffman
University of Texas

Joyce Hood
University of Iowa

Michael Kamil
University of Illinois at Chicago

Lea McGee
Louisiana State University

Peter Mosenthal
Syracuse University

P. David Pearson
University of Illinois

Taffy Raphael
University of Utah

Lenore Ringler
New York University

Diane Schallert
University of Texas

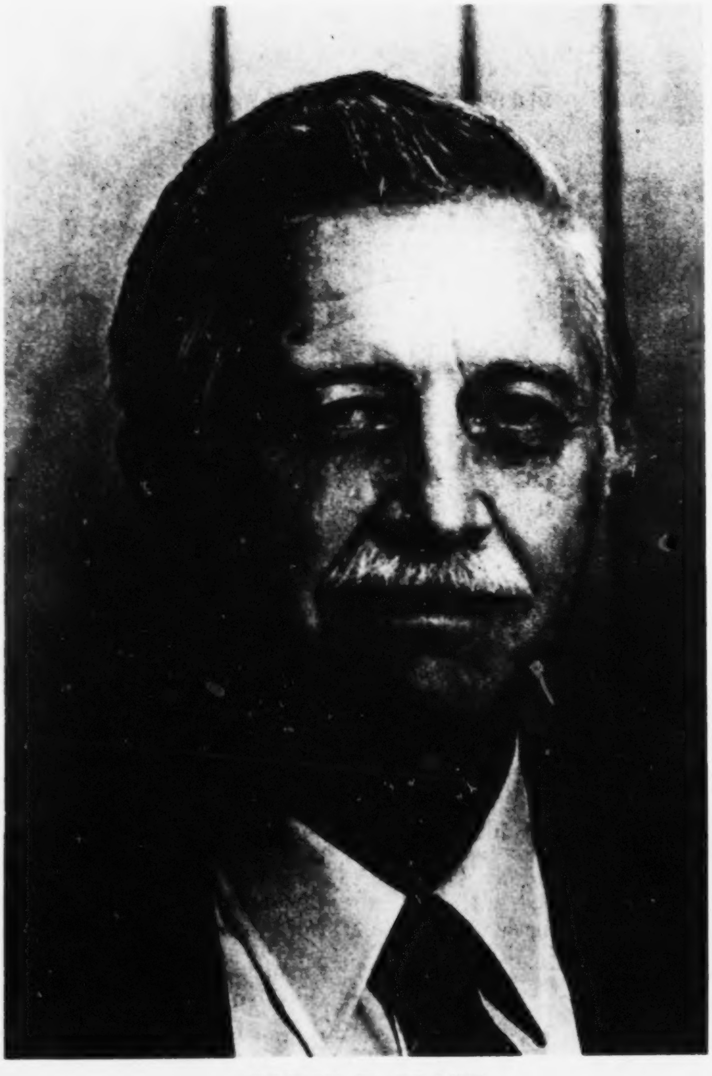
Harry Singer
University of California at Riverside

Dixie Lee Spiegel
University of North Carolina

Elizabeth Sulzby
Northwestern University at Illinois

Barbara Taylor
University of Minnesota

Robert Tierney
Harvard University



GEORGE R. KLARE

Oscar Causey Award

The Oscar S. Causey Award is given annually by NRC to recognize distinguished scholarship and service in the field of reading in general and to the NRC in particular over an extended period of time. The 1981 recipient of the Award is Dr. George R. Klare.

Dr. Klare currently holds the position of Distinguished Professor at Ohio University where he has served over the years in various posts including Department Chairman, Acting Dean, and Dean. His undergraduate and graduate training were completed at the University of Minnesota in the fields of psychology, educational psychology and English. He served in the Army during World War II as a navigator of heavy bombers and received the Purple Heart, as well as the Air Medal for distinguished performance.

As a scholar, Dr. Klare is widely known for his work in the areas of readability, memory, models of communication, and cloze exercises. His well-known statistics book is but one of dozens of publications he has authored in a wide range of fields. He has served as a consultant to numerous agencies including AT and T, Time/Life, Bell Laboratories, the U.S. Navy, Westinghouse, and Western Electric. He has lectured throughout the United States and Europe and served on the editorial board of many professional journals including the *Journal of Reading Behavior*, *Information Design Journal*, and the *Reading Research Quarterly*. As a researcher he has held contracts with the U.S. Air Force, HUMRRO, USOE, Bell Laboratories. He also received a Fulbright travel grant to do research at Open University.

George R. Klare richly deserves the honor of joining our other esteemed colleagues who have received the Oscar S. Causey Award.

TABLE OF CONTENTS

RESEARCH ON READING INSTRUCTION

Applications of Experimental Design Principles to Research on Reading Instruction . . . 1
R. Calfee
An Analysis of the Instruction in Reading Instructional Research . . . 14
G. Duffy and L. Roehler

READING COMPREHENSION: LEARNING AND INSTRUCTION

Computer-Aided Glossing and Facilitated Learning in Prose Recall . . . 24
P. Blohm
Effects of Initial Mention and Orienting Task on Readers' Comprehension of Main Idea . . . 29
S. Belmore, P. Matthews, C. Bridge, S. Moskow and S. Cohen
Reading Comprehension and Attitude of High and Low Imagers: A Comparison . . . 37
E. Cramer
Generating Interactions Between Schemata and Text . . . 42
J. Cunningham
The Skills of Studying . . . 48
T. Estes
Instruction in Inferential Comprehension: An Extension and A Summary . . . 54
J. Hansen and G. Ahlfors
Effects of Cued Attention to Illustrations in Text . . . 60
D. Hayes and J. Readence
The Influence of Metacognitive Knowledge of Expository Text Structure on Discourse Recall . . . 64
L. McGee
Written Questions as Reading Aids in the Middle Grades: A Review of Research . . . 71
D. Memory
Original Versus Adapted Short Stories: The Perceptions of Adolescent Readers . . . 77
J. Mitchell, J. Bradley and W. Ames
Allocation of Instructional Time to Reading Comprehension and Study Skills in Intermediate Grade Social Studies Classrooms . . . 81
A. Neilsen, B. Rennie and A. Connell
Three Studies of Generative Reading Comprehension . . . 85
M. Wittrock

COMPREHENSION: VOCABULARY

The Effects of Context Clue Type and Variations in Content on the Comprehension of Unknown Words . . . 89
B. Carroll and P. Drum
Varied Approaches to Preteaching Vocabulary . . . 94
J. Vaughan, G. Castle, K. Gilbert and M. Love
The Reading Vocabularies of Primary Grade Children of Varying Geographic and Social Backgrounds . . . 99
M. Graves, G. Brunetti and W. Slater

The Effects of Directionality and Proximity of Context Clues on the Comprehension of Unknown Words 105
J. Madison, B. Carroll and P. Drum
Effects of Vocabulary Instruction on Reading Comprehension . . . 110
N. Roser and C. Juel

BEGINNING READING

Teaching Sight Vocabulary through Patterned Language Materials 119
C. Bridge and B. Burton
Evidence of Planning in Dialogue and Monologue by Five-Year-Old Emergent Readers 124
B. Cox and E. Sulzby
Induced Mental Imagery and the Text Prediction Performance of First and Third Graders 131
L. Gambrell
Reading Comprehension and Picture Usage: A Study with First-Graders 136
P. Grinnell
Students' Beliefs and Attitudes About Oral Reading Instruction . 140
J. Hoffman, L. Kastler, M. Nash and J. Daly
The Effects of Differentiated Patterns of Verbal Feedback to Miscues on Word Identification Strategies and Success 145
J. Hoffman, S. O'Neal and R. Clements
The Growth of Letter-Sound Correspondence Knowledge in First Grade and Its Relation to Reading Achievement and Programs . 152
C. Juel and D. Roper/Schneider
The Substrata-Factor Theory of Reading: Differential Development of Subsystems Underlying Reading Comprehension in the First Year of Instruction 157
I. Katz and H. Singer
A Study of Word Errors Produced by Second, Third and Fourth Grade Pupils Reading the Dolch Word List 166
R. Leibert
Judging the Emergent Reading Abilities of Kindergarten Children . 170
B. Otto and E. Sulzby
A Comparison of the Complexity of Children's Dictation and Instructional Reading Materials 177
M. Sampson

RESEARCH AND MEASUREMENT CONCERNS

When Are Qualitative Research Methods Useful in Reading-Related Inquiry? 180
P. Alexander
A System for the Analysis of Word and Context-Based Factors in Reading 185
J. Goldsmith, M. Nicolich and E. Haupt
Classroom Observation Systems: An Historical Perspective 191
J. Logan
Test Preferences and Competencies of Field Educators 196
R. Ruddell and C. Kinzer

Development and Measurement Issues Associated With Reading Classroom Observation Systems 200
W. Rupley and N. Mangano
The Sensitivity of Cloze to Passage Organization 204
T. Shanahan and M. Kamil
Preference for Testing and Reporting of Test Results in Basic Skills: Acquisition vs. Ability to Gain Information from Text, Content-Bound and Content-Free 209
H. Singer and M. Dreher
Summarization Strategy for Improving Reading Comprehension . . 215
J. McNeil and L. Donant

TEACHER EFFECTIVENESS

Teachers' Decision Policies About Grouping Students for Reading Instruction 220
H. Borko
Influential Teachers: Characteristics and Classroom Performance . 227
R. Ruddell and M. Haggard
Specification of Promising Teacher Effectiveness Variables for Reading Instruction 232
W. Rupley, T. Blair and B. Wise
Teachers' Self-Perceptions and Reification of Instruction Within Reading Instruction 237
P. Shannon
The Development of Theoretical Orientation to Reading Among Preservice Teachers: Effects of a Professional Training Program . . 242
J. Stansell, R. Moss and C. Robeck

REVIEW OF RESEARCH

Technology and Reading: A Review of Research and Instruction . 251
M. Kamil

STORY STRUCTURE

Story Schemata: Metatextual Aid to Reading and Writing 261
C. Gordon and C. Braun
Young Children's Detection of Incoherence in Narratives 269
N. Hoover
The Reader-Based and Listener-Based Oral Text Processing of a Sixteen-Year-Old Beginning Reader 274
R. Moss and J. Stansell
Conflict and Conflict Structures in Stories Told by Children . . . 281
D. Spiegel and J. Fitzgerald

WRITING

The Author as Reader: A Comparison of Proficient and Less Proficient College Writers 287
M. Atwell
Semantic Analysis of Children's Writing 294
S. Cohen, C. Bridge and P. Winograd
Children's Concepts About Writing at Home and School 300
J. Black and R. Martin

NRC PROGRAM: DALLAS 1981 305

FOREWORD

Over the past several years reading researchers have been urged to direct their attention toward the important issue of how to improve the effectiveness of reading instruction. A review of the last three *Yearbooks* reveals the positive response of researchers to the call for application. This particular collection of papers, which is representative of the quality and substance of all papers submitted for review, reflects an intensified concern for good educational research on instruction. The comments on research methodology, and the studies in comprehension, beginning reading, and teacher effectiveness all demonstrate the growing awareness of and sensitivity to the classroom environment by reading researchers. However, there is also a note of acknowledgement running through most of these papers that these are only initial steps in a long and complex process of understanding reading and reading instruction in the context of the classroom.

During this first year of our editorship for the *Yearbook*, we have sought the help of a number of our colleagues for the difficult and consuming task of reviewing the many papers submitted for consideration. We wish to thank them for their thoughtful comments and prompt action. A special thanks is owed to Michael Kamil for his encouragement, advice and the *Thirtieth Yearbook*, which was an invaluable guide during our rookie season. In addition, Peter Allen's experience and competence protected us from the numerous hidden potholes on the road to publication. Finally, as with every *Yearbook*, we must thank our creative and industrious colleagues who generated the knowledge which is conveyed in this volume. Their cooperation and thinking made our job considerably easier and intellectually worthwhile.

Jerome A. Niles
Larry A. Harris

ROBERT CALFEE
Stanford University

APPLICATIONS OF EXPERIMENTAL DESIGN PRINCIPLES TO RESEARCH ON READING INSTRUCTION

The present hiatus in support for social science research in general and for reading research in particular has its painful aspects, but may also serve some useful purposes. The past two decades have been a busy time for reading researchers—we now have an opportunity to review what has been accomplished during that time, to rethink our goals and methods, and to prepare ourselves for the tasks of the next two decades.

The body of reading research since 1960 is impressive for its size alone, and much has been learned about substance, theory, methodology, and practice. On the other hand, our most determined efforts have proved disappointing in some respects. In spite of considerable time, energy, and money, it appears to many that we have little coherent advice to offer for policy makers or practitioners about "how to teach reading." What, for instance, has been learned from large-scale social experiments on reading instruction, like the First-grade Reading Study (Bond & Dykstra, 1967) and Planned Variation in Follow-through (for a technical history and list of the voluminous documentation, see Villaume & Haney, 1977)? The answers are not too inspiring:

- The "method" does not seem to make much difference.
- Spending time teaching reading is better than spending time doing something else.
- Teachers make a difference, though why and how are not altogether clear.
- Most of the variance in student performance can be predicted by background characteristics.
- Growth in reading (as presently measured) does not depend greatly on program variables.

One can draw several conclusions from these results: The wrong questions were asked. The analyses were insensitive. The studies were poorly implemented. The experimental method was inappropriate. Whatever the truth of the matter—and it is not my intention to critique these particular experiments here—it does seem clear that the payoff from these studies was much less than expected, and it seems equally clear that the government will not undertake similar experiments in the near future.

My goal in this paper is to explore the possibilities for future research inherent in the concept of *experimental design*. In the large-scale experiments conducted up to the present, little use has been made of design concepts—these may have been "experiments," but they were not especially well designed. We tend to think of experimental design primarily as a method to achieve control and to avoid confounding—in fact it may serve us even more as a tool for formulating our research questions, for laying a foundation for sensitive analyses, and for ensuring proper implementation of the treatments. The focus in this paper will *not* be on statistical analysis, but rather on the application of experimental design principles in the conceptualization and planning of a research study.

Defining terms. To avoid misinterpretation, let me be clear about the meaning of "experiment," "design," and "experimental design." First, I will rely on the dictionary sense of an *experiment* as "a trial to confirm or disprove something doubtful; to discover some unknown principle or effect; to test some suggested truth or to demonstrate a known truth." An experiment is a planned exercise, guided by theory, hunch, or previous experience, which has the aim of clarifying our understanding of how the world works. This definition contrasts in two respects with most discussions of social experimentation (e.g., Saxe & Fine, 1981; Leonard & Lowry, 1979). For one thing, it does not necessarily entail active intervention by the researcher; it encompasses the notion of studies experimentally designed around naturally-occurring variations. For another thing, it does not portray an experiment as comparison of an experimental treatment (one or more) with a control condition (one or more); almost always, this is a poor design.

Next, the word *design*—here again the dictionary is helpful: "an outline, sketch or plan, as of the form and structure of a work of art, an edifice, or a machine to be executed or constructed; the organization or structure of formal elements in a work of art." The concept of design incorporates the analytic planfulness that underlies many intellectual tasks—the notion of constructing a coherent, clearly organized structure.

The designer faces several challenges in planning any research study. First is the systematic analysis of the major dimensions or facets that comprise the problem. This analysis may be guided by theoretical concepts of a more or less formal character. Second is the need to achieve an adequate degree of *control* over the situation, so that the data can be interpreted with a minimum of ambiguity. Control includes the avoidance of confounding, the establishment of the limits of the findings, and the minimization of random ("uncontrolled") fluctuations in the data.

Social science research poses several design challenges that are unique in degree and in kind from other kinds of research:

- The number of factors likely to influence any given situation is relatively large—the problems appear complex.
- The amount of theoretical guidance is relatively slight—trustworthy theories are few, and little use is made of theoretical models where these do exist.
- Interactions among factors may occur, sometimes of substantial magnitude; this complication is not a reiteration of the first point above, by the way. People worry a lot about interactions; in fact, "big" interactions are rare.
- There are severe limits on what can be manipulated and controlled, for reasons both practical and ethical. These limits may be more apparent than real, but more frequent use of natural variations might be wise.
- Social systems (unlike cornfields) are strongly hierarchical, hence the need to take into account the various levels of the system. For example, in research on reading instruction, the student is "nested" within the teacher-class unit, which is nested within the school-neighborhood unit, which is nested within the district-community unit, and so on.
- The resources for carrying out social science research, the magnitude and complexity of the phenomena, and the timelines for investigations, all call for designs that are efficient and productive.

Now to the last term in my list—*experimental design*. The procedures that fall collectively under this rubric have the potential to satisfy all the demands listed above. This potential has not been realized in social experiments to date. One reason for this failure is the overly-simplified conception of an experiment that now prevails. Another reason is the skepticism felt by many researchers about the possibilities for implementing complex designs—the oft-stated acknowledgement of the complicated nature of social phenomena has not been matched by a willingness to address those complications in the design of a project. Finally, the package of techniques for experimental design of social science research is scattered across several domains, and there is no integrated treatment that brings the various pieces together into a coherent, design-oriented framework. An ideal package would encompass at least three domains:

- The concepts of *multifactor design* created by R. A. Fisher (1949; Winer, 1971), chiefly found in experimental and quasi-experimental design textbooks and courses.
- The techniques of survey research, especially those for sample selection in complex populations.
- The area of *test theory*, especially the extensions into generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972).

These three domains encompass most of the concepts and techniques essential for social science research, and most investigators have at least a smattering of knowledge about each. Nonetheless, the knowledge lacks coherence. Individuals tend to be expert in one or another domain; the terminology and styles of thought tend to be idiosyncratic; there is a lack of integrating concepts. It is not too surprising, under these circumstances, that the various parts of most social experiments often have a piecemeal relation to one another.

My thesis in this paper is that the principles of experimental design can serve an integrative function across the three domains. Space will not permit totally adequate discussion of this proposal, but in the remainder of the paper, I want (a) to describe the major principles of experimental design (these often get lost in the shuffle in courses which combine both design and statistical analysis; the ideas disappear in the numbers), (b) to show how these principles can be applied to the particular needs of social science research, and (c) to sketch an example of how one might "experimentally design" a study on reading instruction of the First Grade/Follow-through variety.

Basic principles of experimental design. An experimental design has as its starting point a set of *factors*, variables identified by the researcher as having some influence on the phenomenon under study. The design process is next a matter of selecting *levels* for each of the factors, and of planning suitable arrangements of the various factors and levels, almost like the construction of a complex trap for snaring some elusive and complicated creature.

Factors emerge in rather amorphous form as the researcher grapples with a problem. They assume a variety of shapes and sizes, and one of the strengths of experimental design methodology is that it provides the researcher with a toolbox of techniques for shaping up the factors—i.e., for clarifying the analysis of the problem.

For instance, factors can be *fixed*, *random*, or *fuzzy*. A fixed factor is one where, if the investigation were repeated, the same levels would be studied on that factor. Fixed factors are those which are sufficiently well understood, stable, and valued to warrant careful and continued examination. Binary "yes-no" factors are often fixed: Does the classroom teacher have an aide? Is the student a boy or a girl? Does a passage have accompanying pictures? A random factor is one which would change levels over replications. For instance, we pick new groups of schools, teachers, and students in replicating a study—the original investigation is unconcerned with the particular individuals in that study, but views those as a random sample from a much larger domain. Similarly, a comprehension study might include several passages of a certain type which are chosen at random within that type; in replicating the study, a different set of passages would be chosen. Fuzzy factors are relatively fixed, but with some degree of variation. For instance, a researcher interested in the effects of teaching experience classifies teachers as more (10–20 years) or less (3–9 years) experienced. A replication in which the levels are slightly different (5–12 years versus 13–25 years) ought to show about the same pattern of results.

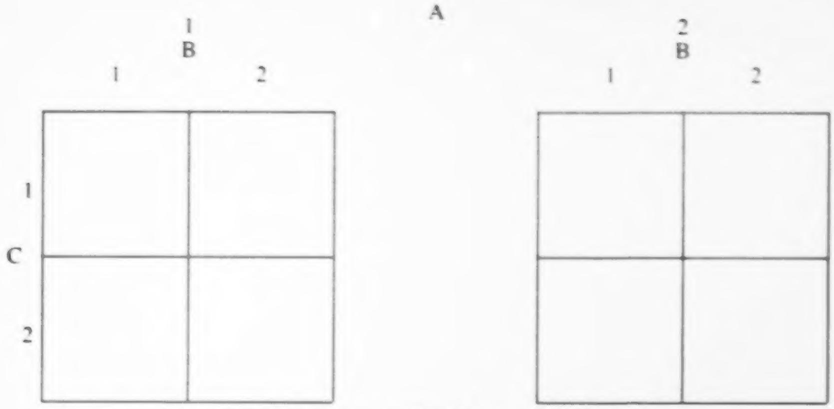
From an experimental design point of view, the important question in each of the above examples is one of interpretation: Is each factor (i.e.; dimension) in a study defined with sufficient clarity that another researcher can replicate the variation in its essential details. A difficulty with many conventional experimental-control designs is that they are *too* fuzzy—too many decisions are wrapped up in the single contrast to ensure that the factor is replicable. By unpacking the various decisions, the investigator becomes clearer about the underlying dimensions in his or her analysis of the problem. A "clean" factor is one in which the levels are distinctive, and therefore easy to replicate—most often there will either be quantitative (how many minutes of reading instruction), or binary (the presence or absence of a certain feature, or contrastive extremes of a hard-to measure phenomenon).

Another important criterion for assessing a set of factors is the degree of independence or *orthogonality* which they permit. Two techniques are used routinely during experimental design to achieve orthogonality: *crossing* and *nesting*. Two or more factors can be crossed when it is possible to create all combinations of them with equal ease. For instance, male and female teachers can be crossed with presence or absence of a classroom aide; it is feasible to assign aides to teachers of either sex. It is more difficult to cross verbal ability and parents' education; the two factors are naturally confounded or correlated, and so two of the combinations (above-median ability with high education, and below-median ability with low education) are more common than the other two. A nested design can handle this situation—in nesting, levels of one factor are defined more or less uniquely within each level of another factor. Thus, the researcher can define the median verbal ability within each level of the parent education factor, with the verbal ability factor now becoming "relative verbal ability within parents' education." A second example is shown in Table 1 for an investigation of the effects of supplementary instruction on vocabulary or comprehension. In this example, the second factor is completely nested; it is defined in totally different ways within each level of the first factor.

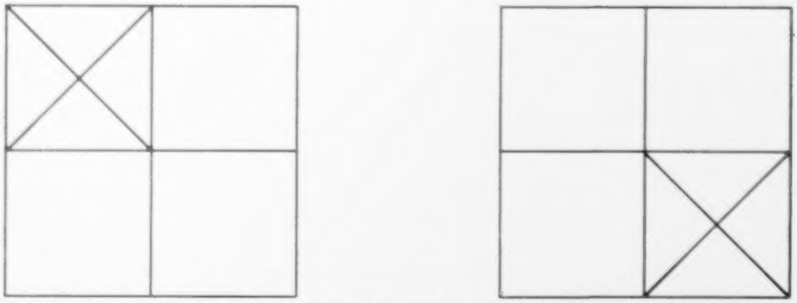
Figure 1 provides a different perspective on the concept of orthogonality. All three designs entail three two-level factors. The top panel shows the eight cells

TABLE 1
Fixed Effects in a Nested Design

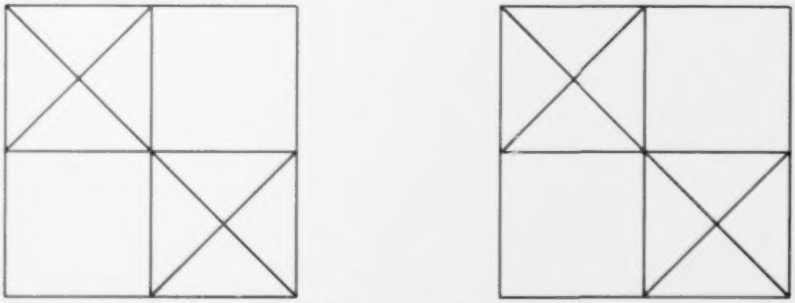
	Factor A: Supplementary Instruction in:	
	Vocabulary	**Comprehension**
Factor B: Type of Instruction	Glossary versus Associates	Narrative versus Exposition



COMPLETE FACTORIAL DESIGN



RANDOM ELIMINATION OF TWO CELLS FROM COMPLETE DESIGN



ONE-HALF FRACTIONAL FACTORIAL DESIGN

FIGURE 1. Orthogonality in Complete and Partial Factorial Designs.

of a completely crossed factorial design. Each of the main effects is completely balanced with the other two factors. For instance, you can see immediately that both levels of B and C are represented equally often at levels 1 and 2 of the A factor, so that any difference between A1 and A2 cannot be attributed to the B or C factors (I am assuming equal numbers of observations in each cell of the design). With a little additional effort, you can confirm that B is balanced with respect to A and C, and that C is balanced with respect to A and B. The interactions are also orthogonal to the main effects and to one another, but this is incidental to the main point being made.

Now look at the middle panel, where two of the eight combinations have been more or less randomly eliminated. At first glance the situation may appear quite satisfactory—the two levels of each factor always occur three times, and so the factors might seem to be equally represented in the design. In fact, they are not. When calculating the mean for level 1 of the A factor, you are averaging over one occurrence of B1 and C1, and over two occurrences of B2 and C2. The situation is reversed when calculating the mean for A2. If either the B and C factor makes a difference, it will contaminate the estimates of the effect of the A factor. The same contamination or confounding occurs in the estimates of the B and C effects. The design is "out of balance"; it is not orthogonal.

In the bottom panel, four of the eight cells have been eliminated, but quite systematically and by no means at random. This reduced design, known as a *fractional factorial* (Cochran & Cox, 1957, Ch. 6A), is arranged to ensure that the factors are each balanced with one another. If you look carefully, you will see that in calculating the mean for A1, you average equally often over the levels of B and C; the same holds when you calculate the mean for A2, and for the effects of the B and C factors. Even though half the design is "missing," the plan is completely balanced. A word of caution—the main effects in this design are confounded with the two-way interactions, and the researcher must assume that these interactions are negligible. These constraints are rather severe in a three-factor design, which provides the simplest example of this approach. In a design with dozens of factors, the main effects are confounded with fourth-order and higher interactions, which can easily be disregarded.

The reader may feel uneasy at this point. Implicit throughout the paper has been the notion of multifactor designs, in which large numbers of factors are cleverly arranged in carefully balanced orthogonal arrangements—and now comes the introduction of incomplete designs with many cells missing. Doesn't this complicate life? The answer is "Not really, and it's worth it."

The benefits: A carefully designed study is more informative, it is better controlled, and the results are more interpretable. A multifactor experiment is a highly efficient way of gathering information, and is essential for assessing interactions. Multifactor designs, when the factors control for variability in the data, can sharply reduce the error of measurement, so that the treatment effects stand out more clearly. The experimenter may choose to ignore factors that are not directly relevant, but the variance associated with such factors does not vanish because of lack of interest; it inflates the error variance, making it harder to see systematic patterns in the data. Finally, with better design comes a clearer understanding of the structure of a problem, and better guidance for future work.

The costs: With so many facets to control, it may seem difficult to handle the design as a whole. This problem is not as serious as it may appear. Experimental design techniques have the virtue of modularity. Rather like a Lego set, one can divide a completed problem into a set of subproblems, plan the arrangements within each set, with the assurance that the parts will fit together as a whole, largely because of the orthogonality of each part.

A second apparent cost is size. A complete factorial design with k two-level factors contains 2^k cells—two factors have 2^2 or 4 cells, three factors have 2^3 or 8 cells, four factors have 2^4 or 16 cells, and so on. The number of combinations increases exponentially, which means very rapidly—each additional factor doubles the size.

Then there is sample size—how many observations are required for each cell in the design? Word of mouth suggests that a sample size of 20 is safe, and that any number less than 10 may be dangerously small. With 10 factors, a complete factorial design requires more than a thousand cells, and with 10 observations per cell, the total number of data points exceeds ten thousand—clearly an intolerable number.

These "costs" arise because of unexamined assumptions underlying common practice. One assumption is that complete factorial designs are better and safer than incomplete designs. In fact, they are not generally easier to implement, and they are tremendously wasteful—in the case of a 2^{10} design, more than 95 percent of the effort goes into gathering information about three-way and higher-order interactions—relatively few researchers would knowingly spend 95 percent of their life exploring such complexities. Fractional designs concentrate the researcher's efforts on questions of major interest (main effects and low-order interactions) and direct time and energy toward complex interactions only when these are of specific interest.

The role of sample size is badly understood by investigators. Whenever an average is to be calculated, sample size is important, but the critical matter is the total number of observations entering into that calculation—*regardless of where the observations come from*. Suppose an investigator has collected 32 observations, half under condition A1 and half under A2. The sample size is a respectable 16 for each of the two means, whether the design has one factor (A) with 16 observations per cell, or two factors (A and B) with 8 observations per cell, or three factors (A, B, and C) with 4 observations per cell, and so on. The means for A1 and A2 can be estimated with the same degree of accuracy $(\sigma/\sqrt{16})$ in all three designs, except that the uncontrolled variability (σ^2) will probably be smaller in the three-factor design because of the improved design. Only if you need great accuracy for estimating the means of particular combinations (e.g., the A1 B2 A3 cell) does it make sense to plan for 16 observations per cell in a three-factor design.

It often helps to think of research in economic terms. Each observation provides the researcher with one piece of information—one degree of freedom, to be precise. The researcher can spend this degree of freedom in various ways—to estimate the main effect of a two-level factor, to estimate an interaction effect, to estimate a piece of some "error" source. Consider several designs for a study in which we have resources to "buy" a set of 32 observations. A conventional experimental-control design entails a single two-level factor with 16 observations per cell. The design takes one degree of freedom to test the effect of the treatment factor—the other 30 degrees of freedom are spent estimating "error." Suppose a two-level "aptitude" factor is now included in the design—three degrees of freedom are spent answering systematic questions (the treatment effect, the aptitude effect, and the interaction between the two), and 28 degrees of freedom for error.

TABLE 2
Allocation of Resources in a 32-Observation Study

	Design	Number of Main Effects Tested	Number of 2-Way Interactions Tested	Degrees of Freedom for Error
	One-way	1	0	30
Complete factorial	Two-way	2	1	28
	Three-way	3	3	25
	Four-way	4	6	21
	Five-way	5	10	16
Fractional	1/2 six-way	6	10	15
	1/4 seven-way	7	9	15
	1/8 eight-way	8	8	15
	1/16 nine-way	9	0	22
	1/32 ten-way	10	0	21

Note: For fractional designs more complex than a 1/2 replicate, the researcher can trade off the degrees of freedom for the interactions and error in a variety of ways. The numbers in the table are reasonable estimates based on experience. Notice that the sample size for main-effect contrasts are the same for all of the designs, because the total sample size is constant, and all observations are used in each main-effect contrast.

There is not space to consider all the possibilities in detail, but just to emphasize the full range of tradeoffs, Table 2 shows the various ways in which a researcher can "spend the degrees of freedom" in a 32-observation study. I have assumed that the research problem is complex and relatively unexplored, so that main effects and two-way interactions are of primary concern, and that other sources can be disregarded. These assumptions seem close to the prevailing conditions confronting instructional researchers. The message in Table 2 is *not* that everyone should use fractional designs—each of the designs in the table is valuable under certain conditions. The point is that experimental designs can provide the researcher with a wide range of alternatives for allocating resources.

Application of experimental design to the Planned Variation Study. This section covers a specific application of experimental design to our area of study. Recall the earlier classification of the domains of methodology in the field—design of the treatments, selection of the sample, and planning of the instrument package. Though experimental design is generally considered only in relation to the first topic, it has value in all three areas. Moreover, the benefits of modularity now become substantial—in principle, the researcher can lay out the overall dimensions of a problem, use these in tackling separately the details of treatment, sample, and instrument package, and be assured that the three parts will fit together fairly well at the end of the planning process.

The task throughout is to divide a big problem into smaller parts, being careful to search for the right "joints" during the carving process (Calfee, 1981). One set of divisions has been suggested above—treatment, sample, and instruments. The substance of the problem also requires some work. In both the First

Grade and Planned Variation studies, the analysis of reading instruction remained at a primitive level, producing indigestible lumps called "reading programs." Some efforts were made both early and late in these studies to identify coherent dimensions of variation, but these efforts were too little and (by and large) too late. The task of classification is demanding but not impossible. One major facet of reading instruction, for instance, is the way in which the major domains within reading (decoding, vocabulary and comprehension) are defined and presented. A related issue is the balance between the academic and nonacademic outcomes of reading instruction. A second facet is the organization of instruction—how are the materials organized, what is the role of the teacher, and so on. A third facet is the context within which reading instruction goes on—the role of the school, the home, the way in which long-term transitions are handled. These components, even if they are not the right divisions, are starting points for analysis. Nor is it the case that such issues were not discussed in the planning of the two studies—such discussions took place, but had little influence on the actual design.

As a concrete illustration of the application of experimental design concepts to instructional research in reading, let me sketch how the Follow-through Planned Variation Study might have been designed. First a few words of background (the project is described in sources mentioned earlier). The study was an effort to evaluate the relative merits of seventeen different programs for the primary school education of disadvantaged students. It was a multiyear program involving more than 50,000 students in hundreds of classrooms from more than 60 sites. The results created considerable heat and a limited amount of light.

I will not pretend to present in the next few pages a satisfactory replacement design, but let me try to sketch what the process would look like. Three basic segments need attention—the treatment, the sample, and the instrument package. Most attention will be given to the first two components.

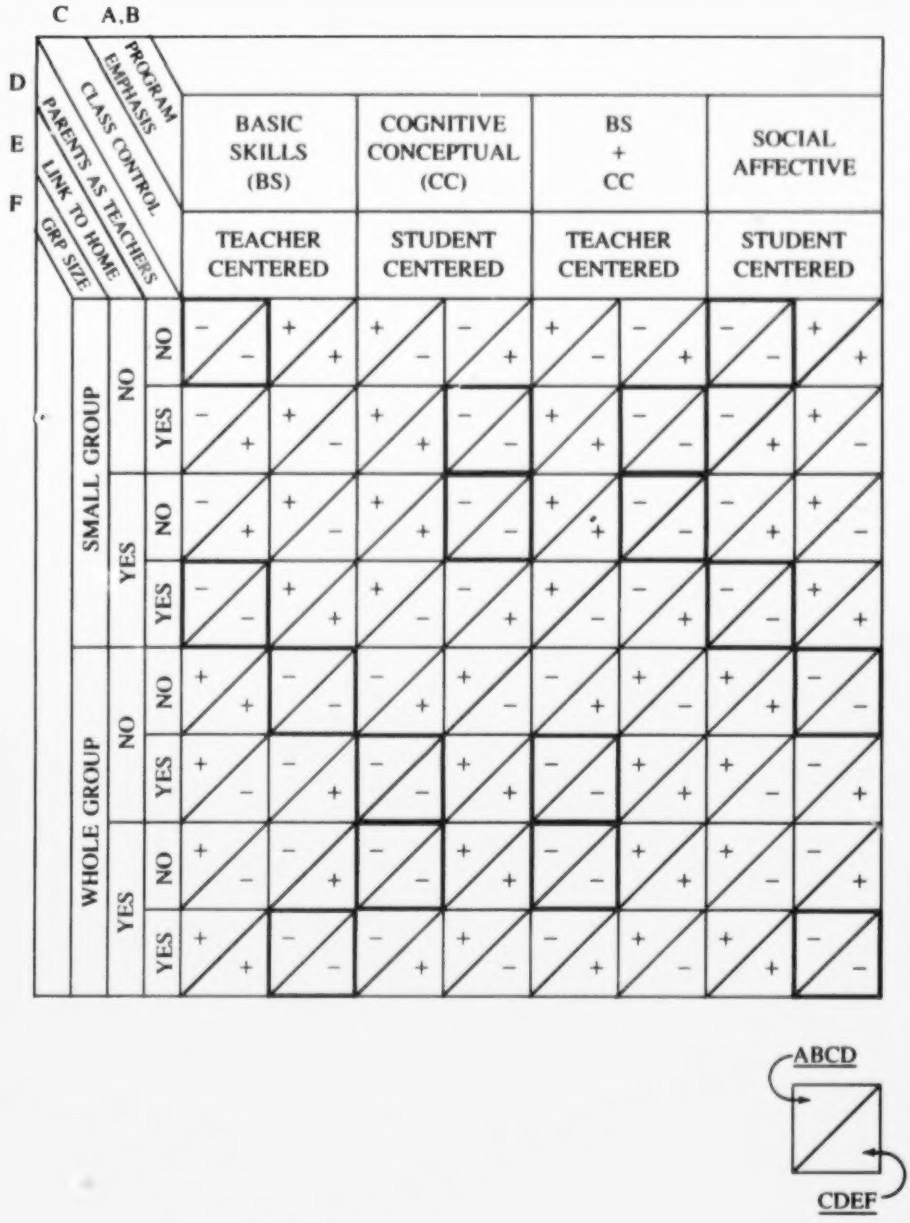
The programs in competition for Follow-through clients varied in several ways—the educational goals, the organization of instruction, and the relation between home and school were the primary dimensions of variation:

- Relative emphasis on (a) basic skills, (b) concepts, (c) both basic skills and concepts, and (d) affective and social outcomes.
- More reliance on (a) teacher-directed or (b) student-centered instruction.
- Most time in (a) whole groups or (b) small groups and individuals.
- Effort to provide linkage to home environment in the classroom—(a) yes or (b) no.
- Effort to use parents to supplement and reinforce school learning—(a) yes or (b) no.

These variations, which can be directly converted to factors in a rough-and-ready fashion, deserve further refinement. Nonetheless, they represent a starting point that is fairly close to the analysis of program variables in Planned Variation. The seventeen programs in that study were more or less random combinations of these variables; each program focused on some factors and ignored others in relatively idiosyncratic combinations.

In Figure 2 is shown an experimental design that provides systematic control over the treatment factors. There are a total of six 2-level factors (the 4-level Emphasis factor can be handled as a pair of 2-level factors), which yield a total of 64 unique programs—too many for practical purposes. The ¼-fractional plan in the figure requires the preparation of 16 programs, comparable in number to the Follow-through study, but designed to allow clear evaluation of

20


C
A,B
D
E
F
PROGRAM EMPHASIS
CLASS CONTROL
PARENTS AS TEACHERS
LINK TO HOME
GRP SIZE
ABCD
CDEF

GRP SIZE	LINK TO HOME	PARENTS AS TEACHERS	BASIC SKILLS (BS)		COGNITIVE CONCEPTUAL (CC)		BS + CC		SOCIAL AFFECTIVE	
			TEACHER CENTERED		STUDENT CENTERED		TEACHER CENTERED		STUDENT CENTERED	
SMALL GROUP	NO	NO	− / −	+ / +	+ / −	− / +	+ / −	− / +	− / −	+ / +
		YES	− / +	+ / −	+ / +	− / −	+ / +	− / −	− / +	+ / −
	YES	NO	− / +	+ / −	+ / +	− / −	+ / +	− / −	− / +	+ / −
		YES	− / −	+ / +	+ / −	− / +	+ / −	− / +	− / −	+ / +
WHOLE GROUP	NO	NO	+ / +	− / −	− / +	+ / −	− / +	+ / −	+ / +	− / −
		YES	+ / −	− / +	− / −	+ / +	− / −	+ / +	+ / −	− / +
	YES	NO	+ / −	− / +	− / −	+ / +	− / −	+ / +	+ / −	− / +
		YES	+ / +	− / −	− / +	+ / −	− / +	+ / −	+ / +	− / −

FIGURE 2. Fractional Plan for Follow-through Treatment Factors.

the underlying factors. The seventeen Planned Variation programs, if they could be fit into the figure, would fail to allow an orthogonal evaluation of the various factors. In fact, several of the programs cannot be classified on certain dimensions—an *experimental design approach leads the researcher to a more* careful analysis of the treatment conditions than is typical when program "labels" are the chief source of variation.

Next is the sample. The Planned Variation sample was selected according to region (Northeast, North Central, South, and West) and community size (Large, Medium, Small, and Rural). Once more these dimensions could stand some refinement, but they are a reasonable starting point. In addition, I would be inclined to take into account the characteristics of the principals and the teachers. Here are some suggested dimensions:

- School principal focuses primarily on (a) business management or (b) instructional leadership.
- Teacher experience is (a) relatively limited or (b) relatively extensive.
- Teacher values and goals are primarily (a) cognitive and academic or (b) social and affective.

The 16 combinations of region and community size, together with the 8 combinations of principals and teachers, comprise a full factorial with 128 cells—let us assume that the sample consists of two classrooms fitting each of these descriptions, one assigned at random to a Planned Variation treatment, and the other serving as a matched control. The total number of classrooms called for in this design is about one-tenth the number in the Follow-through study, as nearly as I can tell. Once the classrooms are selected, design principles can be used to classify the students, and for random selection of target students; this step enhances the design, but poses no additional complications.

We now have two matched cohorts of 128 classrooms, balanced on a variety of dimensions. The next step is to assign the 16 treatment combinations to the 128 Follow-through classrooms. The details are tedious, and will not be illustrated. In effect, the task is to arrange the 128 schools in 16 balanced blocks of 8 each, and then assign one of the 16 programs to the classrooms in each block so as to ensure proper balance between the treatment factors and the classroom factors. Another way to look at this step is to note that the total design has 13 factors (6 treatment factors and 7 sample-selection factors), which must be fit into the 2^7 or 128 classes, a reduction of 1/64. A design of this complexity may seem extraordinary; in fact, the plans available from the National Bureau of Standards (1962) can handle up to 16 factors and a reduction of 1/256 in the size of the design, which is substantially more complex than the one being proposed.

The actual implementation of a program is recognized increasingly as a significant variable—how can this dimension be handled in the design? Let us assume that implementation is to be treated in omnibus fashion, such that the 8 classes within a particular treatment combination are ranked, and then classified as above or below average in that treatment. (The degree of implementation is an outcome in its own right, and the design would permit evaluation of the degree to which certain dimensions of the treatment or the sample promoted or hindered implementation.) This information can then be fed back into the design as a partly nested factor (relative implementation) for analysis of student outcomes.

A few words about the instrument package for assessing student responses to the various programs. In the Planned Variation Study, there were instruments for assessing outcomes in each of the three areas of emphasis—basic skills, conceptual growth, and social-affective reactions. The instruments, mostly paper-and-pencil group-administered, were "off-the-shelf," and differed from one another to such an extent that comparisons across instruments were hard to justify—the tasks and the materials were too dissimilar.

At least two other dimensions could have been included in the design of the instrument package. The first, degree of structure, is directly related to the treatment program. In a test, the student can either be given fairly direct guidance about what is to be done and how to do it, or can be presented with a general problem and given some freedom to tackle it. Multiple-choice tests are relatively structured, by and large (the answers are provided), and thus are a better fit to more structured treatment programs. The advantages of training in independent, conceptually-oriented problem solving are more likely to surface during the performance of tasks that call for such skills.

The second dimension should probably be included in all instructional research—the contrast between the specific materials, tasks and contexts that were used during training and a set of contrastive materials, tasks and contexts selected at random from the larger domain of transfer of training. While there was some variation in the degree to which the Follow-through tests matched the training conditions, for the most part the tests measured transfer—there were no systematic measures of program-related learning.

In conclusion. Cronbach (1971) has argued that determining the validity of a test is not a matter of "proving" that it predicts a particular outcome—validity is instead a continuing process of establishing the limits within which a set of test results may be interpreted in a particular way. Test validation is a set of procedures for extending our knowledge about the nature of the behavioral world. In like manner, one can argue that social science experiments are not for "proving" that one program or treatment is better than another—all such efforts must fail sooner or later. Experiments like Planned Variation in Follow-through are much more valuable when they are designed to improve our knowledge of the workings of complex social systems. To the degree that such studies are planned to contrast "what is" with some new variation of "what might be," to contrapose one complex program against another, or to demonstrate the universal effectiveness of a simplistic solution to a complex problem, social science experiments in education and elsewhere will have little payoff.

Well conceived and properly designed, such studies have considerable payoff, in my opinion. The concepts and procedures of experimental design, while by no means the only tool at our disposal, provide a powerful technology for tackling the problems before us. This potential has yet to be fully realized. The Planned Variation experiment could have provided 10 to 100 times as much information at 1/10 the cost if it had taken full advantage of the techniques of experimental design. Nor need the level of implementation be at the order of magnitude of the Follow-through study to benefit from these procedures. Juel (1977, 1980) has demonstrated the use of factorial design principles to study the complexities of the reading process and reading instruction in what might be viewed as educational micro-environments. We are not likely in the next few years to see investments on the order of magnitude of Planned Variation—nor even one-tenth that magnitude. But instructional research will continue, and it behooves us to learn as much as we can from the data we can muster—effective application of experimental design methods can move us in that direction, enhancing our knowledge and our awareness of how we come to know what we know.

ACKNOWLEDGEMENT

Preparation of this paper was funded in part by: National Institute of Education, Grant No. G-79-0139; Office of Education, Grant No. G007903258; and the Center for Advanced Study in the Behavioral Sciences.

REFERENCES

Bond, G. L., & Dykstra, R. The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, 1967, *2*(4) (complete volume).

Calfee, R. C. Cognitive psychology and educational practice. In D. C. Berliner (Ed.), *Review of Research in Education*. Washington, D.C.: American Educational Research Association, 1981.

Cochran, W. G., & Cox, G. M. *Experimental designs*. New York: John Wiley & Sons, Inc., 1957.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, J. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley & Sons, 1972.

Fisher, R. A. *The design of experiments*. Edinburgh: Oliver & Boyd, Ltd., 1949.

Juel, C. L. *An independent-process model of reading for beginning readers*. Unpublished doctoral dissertation, Stanford University, 1977.

Juel, C. L. Comparison of word identification strategies with varying context, word type, and reader skill. *Reading Research Quarterly*, 1980, *15*, 358–376.

Leonard, W. H., & Lowry, L. F. Was there really an experiment? *Educational Researcher*, 1979, *8*, 4–7.

National Bureau of Standards. *Fractional factorial experimental design at two levels* (Doc. C.13.32.48). Washington, D.C.: U.S. Department of Commerce, Applied Math Series, 1962.

Saxe, L., & Fine, M. *Social experiments: Methods for design and evaluation*. Beverly Hills, CA: Sage Publications, 1981.

Villaume, J., & Haney, W. *The Follow-through planned variation experiment* (Vol. V, appendix; analysis of interim Follow-through evaluation reports). Cambridge, MA: Huron Institute, 1977.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

GERALD G. DUFFY
LAURA R. ROEHLER
Michigan State University

AN ANALYSIS OF THE INSTRUCTION IN READING INSTRUCTIONAL RESEARCH

Reading researchers have long been interested in the processes of reading and how individuals learn to read. Less emphasis has been placed on instruction and what teachers do to teach reading. Recently, however, a change has occurred. Stimulated by findings from research on teaching and by the needs of practitioners, more and more reading research has focused on instructional issues. For instance, Durkin (1978–79) has examined the comprehension instruction in elementary classrooms, Calfee and Piontkowski (1981) have looked at decoding instruction, Hansen (1981) has studied the teaching of inferencing, Cohen and Stover (1981) have examined computerized teaching of fluency.

For those of us who are concerned about the instructional issues associated with reading and reading comprehension, this interest in instruction is a happy circumstance. It is not unreasonable to expect that, as a result of these efforts, we will soon be in a position to inform the practice of teachers and teacher educators regarding effective teaching of reading.

Currently, however, progress toward this goal is slow. This is partly because researchers of reading instruction, having been less than precise in our use of the term "instruction," are now having trouble defining what it is we are studying. For instance, there is a debate over what Durkin really meant when she reported that the teachers she observed spent little time *instructing* comprehension; Hodges (1980) argues that instruction is more than what Durkin said it was while Heap's (Note 1) critique of Durkin is based on yet another definition of what instruction is. There are other examples of instruction meaning something different to different researchers. Yet we use the term as if we share a common definition. As a result, we sometimes have difficulty interpreting the implications of even the finest research studies because we do not know where one person's "instruction" intersects with another's. Such lack of precision makes it difficult for us to talk to each other, and the teachers and teacher educators who are the consumers of our research have difficulty knowing how our findings inform their practice. Hence, we need to think rigorously about what the instruction is in reading instructional research.

This paper is a first step in that direction. It suggests analytic categories for describing and thinking about instruction, particularly that found in current research of comprehension teaching. We have no illusions regarding the long-term utility of our effort. It no doubt will and, indeed, *ought* to be soon made obsolete by more refined, more elegant and more descriptive efforts. It does, however, serve as an example of the type of thinking we feel must be done if we are to become more precise in our study of instruction.

THE ANALYTIC CATEGORIES OF INSTRUCTION

To clarify thinking about the instruction in reading instructional research, we analyzed studies currently being reported and tried to rationally determine what instructional distinctions are being made. The result was eight analytic

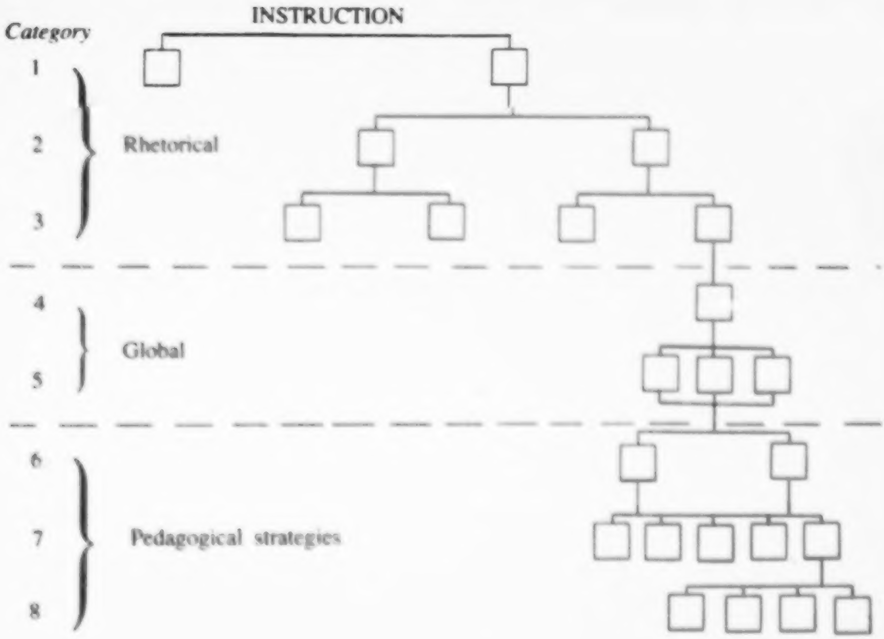


FIGURE 1. Analytic categories to describe reading instructional research.

categories, starting with the broadest distinctions about the meaning of instruction and moving steadily through finer and finer distinctions. Each analytic category is identified with a numeral and includes one or more distinct conceptual types, although we realize that in practice such distinctions are seldom clear and that, indeed, teachers may sometimes deliberately combine types.

Figure 1 provides an overview of the direction the analysis takes. Note that the eight categories are divided into three sub-sets: Categories 1–3 describe the rhetorical distinctions that are often made about instruction (see Figure 2); Categories 4 and 5 describe the global ingredients which are descriptive of the instructional modes currently being investigated (see Figure 3); and Categories 6–8 describe the specific pedagogical behaviors being studied in reading instructional research (see Figure 4). Note also that the analysis progresses to the right in Figure 1, rather than developing in balance. This is because the reading instructional research currently being reported is almost exclusively of the "direct instruction" mode. Because the other modes are discussed but seldom researched, there was no instruction to analyze in these modes.

Categories 1–3: Rhetorical Distinctions

The first three categories are rhetorical: they are talked about but seldom researched. They focus on broad distinctions and result in four theoretically-feasible modes. While only one of these modes is represented in current reading instructional research, all four are important since many of the debates about the way comprehension ought to be taught are based on these rhetorical distinctions.

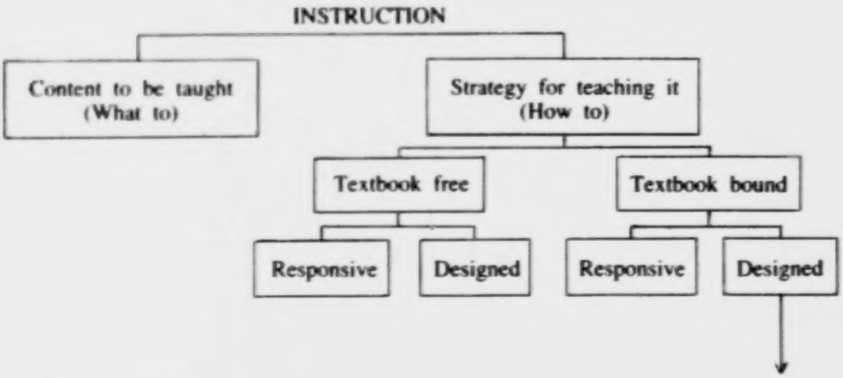


FIGURE 2. Rhetorical distinctions.

Category 1. The first category is strictly definitional. It makes a distinction between what to teach and how to teach it. The question of "what to teach" is a curriculum issue; its focus is the content to be selected for instruction. Instruction, however, refers to how one imparts this content to learners. While instruction cannot occur without content, and teaching effectiveness depends both on teaching the right thing and teaching it well, the "what to" and "how to" questions are nevertheless conceptually different. Concerns such as whether we should include more writing in the reading period, or whether we should teach syllables rather than phonograms are problems of "what to teach." Instruction—the "how to"—occurs after such content decisions have been made.

Category 2. The second category focuses on the source of the materials of instruction, since all reading instruction implies that something must be read. Two conceptual types are possible here (although we remind you again that, in practice, they can be combined). The first conceptual type is "textbook-bound" instruction, and is based on the assumption that students will read a textbook of one kind or another. The basal textbook is the prevalent example found at the elementary level, although softback programmed materials, instructional kits and even computer assisted instruction are of this conceptual type. The second is "textbook free" reading instruction, and is based on the assumption that students will not use a textbook; instead, reading materials will be selected and/or created by the students themselves. Most versions of the language experience approach are typical of this type, as are "individualized reading" approaches which call for the use of trade books.

Category 3. The third category focuses on the teacher's instructional role when bringing students together with the materials. This role can be one of two types: in the first, the teacher follows the directives of an instructional design (either prescribed for or developed by the teacher) which includes a logically developed sequence of both content and instructional activities; in the second, the teacher allows children to explore and experience the materials in an unstructured manner and responds in appropriate ways to the "teachable moments" which arise from such encounters. This distinction between "designed" and "responsive" instruction is crucially important. Designed instruction provides op-

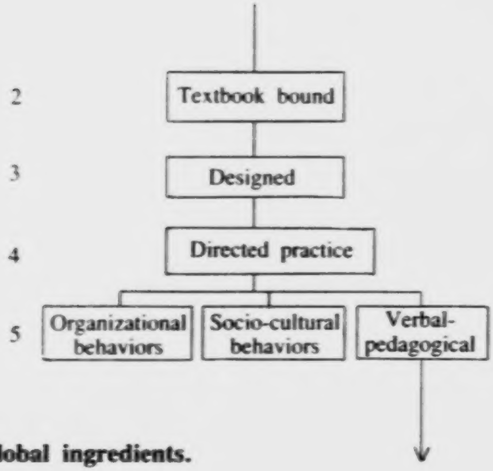


FIGURE 3. Global ingredients.

portunity for students to learn; responsive instruction waits for pupil responses which in turn provide the teacher with the opportunity to teach. Designed instruction expects the teacher to adhere closely to the instructional prescription; responsive instruction expects spontaneous and inventive teacher responses to children's interactions with text. Designed instruction assumes that instruction *creates* learning outcomes; responsive instruction assumes that instruction *facilitates* learning outcomes. In short, designed instruction is what most people are talking about when they describe "direct instruction" in reading, while responsive instruction is more consistent with the practices espoused by "open classroom" advocates.

Summary. As seen in Figure 2, instructional diversity is theoretically possible. However, only the textbook-bound/designed mode of instruction is reflected in the reading instructional research currently being reported. Consequently, we now pursue our analysis in that direction.

Categories 4 and 5: Global Ingredients

Currently reported reading instructional research has two characteristics. First, as noted above, nearly all the studies reflect the textbook-bound/designed mode of instruction. Second, within this mode, certain instructional characteristics prevail. These are discussed in Categories 4 and 5.

Category 4. The fourth analytic category for describing and thinking about instruction is the only one for which there is no apparent disagreement. Regardless of one's instructional position, it is agreed that instruction must solicit a response from students and that such responses are best obtained through directed practice. In short, practice is a necessary part of instruction. Even advocates of textbook-free and responsive modes of reading instruction agree that practice is important (although they prefer to call it "experiences with text" or "encounters with print").

Category 5. The complexity of instruction is revealed in the fact that, while practice is central to instruction, a variety of teacher instructional behavior is expected during practice. The research includes three distinct types.

We label the first "organizational." These are the instructional behaviors teachers use to organize their classroom, including the scheduling and management behaviors associated with allocated instructional time and engaged time on task, grouping behaviors, and pacing behaviors. Many of the findings from research on teaching, especially the process-product studies, have focused attention on the importance of such organizational behaviors.

The second type is the teacher's socio-cultural instructional behaviors. Here, the focus becomes the teacher's role in forging working relationships with students, in maintaining activity flow, in resolving various role conflicts and in establishing and maintaining the classroom social system. The work of McDermott (1977), Mehan (1979), Doyle (1979), Au and Mason (1981) and others are particularly relevant here. The common focus is the teacher's need to solicit student responses within an environment governed by social and cultural conventions.

The third type is labeled "verbal-pedagogical." It is universally assumed that teachers will say something to students when soliciting responses during practice and that what they say is pedagogically based—that is, that teachers' verbal interactions with students embody pedagogical strategies which increase the chances that the student will respond successfully. The various kinds of verbal-pedagogical strategies found in current reading instructional research are discussed in Category 7.

Summary. As illustrated in Figure 3, the textbook-bound and designed instruction calls for directed practice and teacher behaviors associated with organization, social management and verbal-pedagogical interactions. The latter is the focus of categories 6–8.

Categories 6–8: Specific Pedagogical Behaviors

As noted above, effective teaching requires (1) that classrooms be organized, (2) that instruction be conducted within a social system and that teachers say helpful things to students. The specifics of the latter is our focus now.

Category 6. It is generally agreed that teachers ought to say helpful things to students. How the teacher knows what to say is less clear. Two positions prevail. The first (and most prevalent) is typified by the standard basal textbook; the design prescribes both the curricular sequence and the sequence of instructional activities but stops short of specifying the exact words the teacher must use when presenting instruction. The second, exemplified by Distar, also prescribes a design for curricular sequence and instructional activities, but adds a script for teachers to follow verbatim when presenting instruction. The issue here is whether instructional communications are to be created and articulated by teachers themselves or created and scripted *for* teachers by "master developers" (Rosenshine, Note 2).

Category 7. Whether scripted or non-scripted, instructional communication in current instructional research includes five types: setting expectancies, giving directions, monitoring with corrective feedback, modeling and explaining. At first glance, these appear to be distinctly different. However, as we have argued elsewhere (Duffy and Roehler, 1982), this is an illusion. There are actually only two distinct conceptual types of verbal-pedagogical behavior.

This first is based on the assumption that by exposing students to a task they will spontaneously develop an intuitive understanding of how to do it. Such a "spontaneous generation" view of instruction includes the first four verbal-pedagogical behaviors. For instance, the pedagogical strategy for "setting ex-

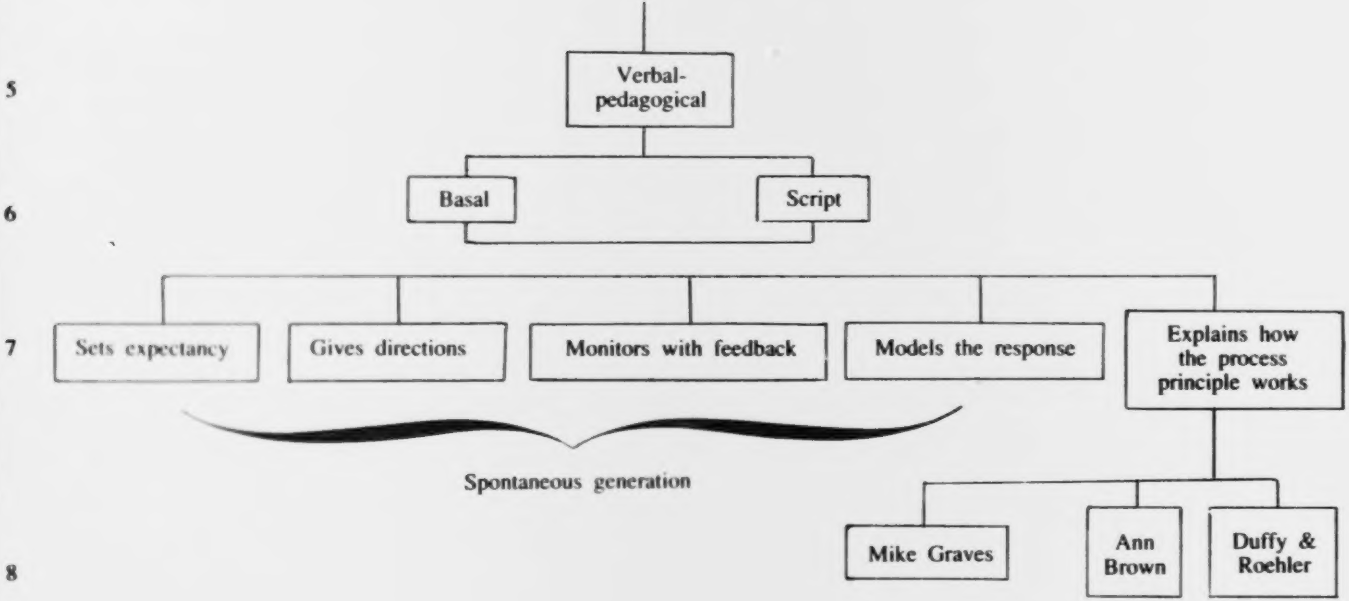

5
Verbal-pedagogical
6
Basal
Script
7
Sets expectancy
Gives directions
Monitors with feedback
Models the response
Explains how the process principle works
Spontaneous generation
8
Mike Graves
Ann Brown
Duffy & Roehler

FIGURE 4. Pedagogical strategies.

pectancies" is that teachers can help children learn by establishing an instructional environment which calls for the use of the desired outcome. A recent research example is Hansen's (1981) question strategy, in which she asked only inference questions of one group on the asssumption that, by posing only such questions, the student is encouraged to learn how to answer them. In short, students will respond to the teacher's expectancy by spontaneously developing an understanding of how to do the cognitive processing required.

A similar belief is reflected in giving procedural directions. The pedagogical strategy is that the student, if given clear directions for *what* to do, will learn *how* to do it. A recent example is the directions given to the subjects in the Carver and Hoffman (1981) study of computer-assisted instruction of fluency.

In monitoring with corrective feedback, the teacher monitors pupil responses to practice and supplies correctives when errors are made. Such an approach is cited by Cohen and Stover (1981) in their study of comprehension of math story problems, by Hodges (1980) in her ciritque of Durkin and by Heap (Note 1) in *his* critique of Durkin. Similarly, Allington's (1980) studies of teacher interruption behavior and the Hoffman, O'Neal and Baker (Note 3) study of teacher feedback to miscues reflect this view. The pedagogical strategy is that teachers can help children learn if, after errors are made, corrections are provided. While much depends here upon *how* corrections are provided, the assumption frequently seems to be that, if one is corrected enough times, the understanding of how to do the task in the future will spontaneously appear.

Modeling can also be "spontaneous generation" if the teacher demonstrates only the response, but not *how* the response was obtained. Such modeling is emphasized in Carnine and Silbert (1979) and in Hansen's (1981) Strategy treatment in which she modeled the inferencing response using a weaving analogy without demonstrating the cognitive processing employed. The assumption is that an intuitive understanding of how to do the *processing* will be achieved spontaneously by virtue of repeated exposure to a model of the *task*.

Only one type of teacher verbal-pedagogical behavior stands in contrast to "spontaneous generation." It is teacher explanation behavior. Based on the assumption that teachers should verbally explain to students not only what the correct response is but also the processing involved, it focuses on (1) identifying the cognitive processing which readers use when successfully completing the task and (2) explicitly explaining to students how to do such processing. It reflects a belief that, because language is a system of conventions and principles which we all share (albeit sometimes intuitively and unconsciously), students who do not spontaneously grasp these conventions and principles should be made aware of them through explicit explanation. Examples of various approaches to teacher explanation behavior follow.

Category 8. There are three examples of teacher explanation. The first is the work of Brown and her colleagues (1981) at the Center for the Study of Reading at the University of Illinois. They have conducted research of instruction in which students are made consciously aware of their own mental processing so that they can make sense out of the task and monitor their own thinking. As Brown (1976) states, it is a process of ". . . externalizing an internal mental event." It is startingly different than "spontaneous generation" behavior because it does *not* assume that all children learn as a result of repeated exposure to the task but, rather, that some profit from explanations which make explicit the processes of reading and thinking.

Graves (Note 4), working on the teaching of vocabulary meaning, also emphasizes teacher explanation behavior. When teaching prefixes, for instance, he makes students conscious of both the facts about prefixes (the conventions and principles which govern their use) and the strategies for using prefixes (the mental processing one uses when unlocking prefixed words) and uses rehearsals as a means for making the thinking process explicit.

Working within the framework of a teacher's daily classroom practice, we also emphasize teacher explanation (Duffy and Roehler, Note 5). Instruction focuses on explicitly stating the purpose of the lesson, why it is important and the convention or principle to be employed; (2) modeling the *thinking* involved in completing the task with the intention of externalizing the internal processes employed; (3) providing a series of assisted steps which are gradually reduced until the child can do the task alone; and (4) checking to make sure that the students are consciously aware of what has been taught, why it is useful and the secret to being able to do it successfully. The intent is not only to teach students to get the right answer but to also make them consciously aware of *how* they get it.

Summary. Two broad approaches to instructional communication exist. Reading instructional research, however, is dominated by "spontaneous generation." Teacher explanation, which assumes that some students will not be sensitive to the implicit cues of language and will need teachers to explicitly surface the internal mental processing for them, is rarely studied. Clearly, however, these are two conceptually different ways to think about reading comprehension instruction.

CONCLUSION

It is always difficult to design rigorous research studies and to report findings in ways which communicate totally. When the object of study is as complex as instruction in reading, miscommunication occurs and progress is slowed.

Consequently, it is helpful to analyze instruction in terms of categories and to examine the distinctions between and within categories. By doing so, three benefits result. First, it provides a glimpse of the instructional gestalt, and helps us place individual studies within this gestalt. Second, it helps us know what we are getting smarter about. For instance, we are getting smarter about textbook-bound and designed reading instruction (direct instruction) but we are not learning much about other modes of instruction. Third, it helps us understand related issues. For instance, regarding the debate about instructional decision-making in reading (Borko, Shavelson, & Stern, 1981; Duffy, 1982), one's expectations about what constitutes instructional decision-making varies depending upon one's referrent for "instruction" with teachers employing textbook-free and responsive modes doing a different kind of decision-making than teachers using a standard basal. Most important, however, it makes for clarity of communication. Researchers are better able to talk to each other about their work, and teachers and teacher educators understand better how the research informs practice.

The task is not done, however. If researchers are to become truly precise in the design of reading instructional research, we must ultimately possess a "grammar of instruction." We hope this paper is a step toward achieving this goal.

REFERENCE NOTES

1. Heap J. *Understanding classroom events: A critique of Durkin, with an alternative.* Paper presented at the State of the Art Conference, State University of New York-Albany, November, 1981.
2. Rosenshine, B. *Meta-analysis of process-product research.* A paper presented at the Invisible College of Researchers on Teaching, UCLA, April, 1981.
3. Hoffman, J., O'Neal, S., & Baker, C. *A comparison of inservice and preservice teacher verbal feedback to student miscues across two difficulty levels of text.* A paper presented at the National Reading Conference, San Diego, November, 1980.
4. Graves, M. *Teaching prefixes.* Unpublished paper, University of Minnesota, 1981.
5. Duffy, G., & Roehler, L. *An exploratory study of teacher explanation behavior.* A paper presented at the annual meeting of the International Reading Association, Chicago, April, 1982.

ACKNOWLEDGEMENTS

Most of the thinking for this paper was done while the authors were on sabbatical leave from Michigan State University during 1980–1981. The writing was done under the auspices of the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Teaching Division of the National Institute of Education, United States Department of Education. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-76-0073)

The authors gratefully acknowledge the tangible assistance provided by their colleagues at Michigan State University, particularly Drs. Linda Patriarca, George Sherman, Lee Shulman and Roy Wesselman.

REFERENCES

Allington, R. Teacher interruption behavior during primary grade oral reading. *Journal of Educational Psychology*, 1980, *72*, 371–377.

Au, K. H., & Mason, J. M. Social organization factors in learning to read: The balance of rights hypothesis. *Reading Research Quarterly*, 1981, *17*, (1), 115–152.

Borko, H., Shavelson, R., & Stern, P. Teacher's decisions in the planning of reading instruction. *Reading Research Quarterly*, 1981, *16*, 449–460.

Brown, A., Campione, J., & Day, J. Learning to learn: On training students to learn from texts. *Educational Researcher*, 1981, *10* (2), 14–24.

Brown, A. Semantic integration in children's reconstruction of narrative sequences. *Cognitive Psychology*, 1976, *8*, 247–262.

Calfee, R., & Piontkowski, D. The reading diary: Acquisition of decoding. *Reading Research Quarterly*, 1981, *16*, 346–373.

Carnine, D., & Silbert, J. *Direct instruction reading.* Columbus, Ohio: Charles Merrill, 1979.

Carver, R., & Hoffman, J. The effect of practice through repeated reading on gain in reading ability using a computer-based instructional system. *Reading Research Quarterly*, 1981, *16*, 374–390.

Cohen, S. A., & Stover, G. Effects of teaching sixth-grade students to modify format variables of math word problems. *Reading Research Quarterly*, 1981, *16*, 175–200.

Doyle, W. "Classroom Tasks and Students' Abilities." *Research on teaching: Concepts, findings, and implications.* (P. Peterson and H. Walberg, eds.). Berkeley: McCutchen Publishing, 1979.

Duffy, G. Commentary: There's more to instructional decision-making in reading than the "empty classroom." *Reading Research Quarterly*, *17* (1), 1982.

Duffy, G., & Roehler, L. Commentary: The illusion of instruction. *Reading Research Quarterly*, *17* (2), 1982.

Durkin, D. What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 1979, *14*, 481–533.

Hansen, J. The effects of inference training and practice on young children's reading comprehension. *Reading Research Quarterly*, 1981, *16*, 391–417.

Hodges, C. Commentary: Toward a broader definition of comprehension instruction. *Reading Research Quarterly*, 1980, *15* (2), 299–306.

McDermott, R. P. Social relations as contexts for learning in school. *Harvard Educational Review*, 1977, *47* (2), 198–213.

Mehan, J. *Learning lessons: Social organizations in the classroom*. Cambridge: Harvard University Press, 1979.

PAUL J. BLOHM
University of Southwestern Louisiana

COMPUTER-AIDED GLOSSING AND FACILITATED LEARNING IN PROSE RECALL

As computers enter the mainstream of educational instruction in the 1980s, teachers will be searching for courseware which offer more than drill-and-practice of specifically taught skills. This investigation was devised to initiate a line of research into the compensatory effects of a computer-based system of reading aids for increasing learner's comprehension and retention of text. The purpose of this study was to determine whether the computer-aided gloss paraphrase (a type of reader-activated superimposed intratext notation that replaces abstractly-composed text content with concrete referents to promote readers' cognition) could be used to facilitate students' recall ability. Results of this experiment and related studies might suggest a viable procedure for designing computer programs which can be in content-area subjects to develop and guide text comprehension skills and strategies.

A recent instructional research effort which has synthesized the theoretical aspects of the interactive reading model, schema theory and metacognition to improve readers' application of skills for comprehending text has been to gloss content-area subject matter. *Gloss* is a generic term which refers to marginal or intratext notations which are written to actively direct readers' attention to segments of text where specific skills and strategies can be applied to comprehend the information. Results of pilot studies with prototype gloss materials and procedures have indicated a practical and efficient means for clarifying, systematizing, and refining the glossing technique for use in the content classroom. (See Otto, White, & Camperell, 1980, for detailed results of pilot studies and subsequent investigations and for examples of the format and content of "glossed" text.) Glossing, in effect, could provide the teacher with a practical means for helping the student when the two are away from each other to both promote the student's acquisition of the assigned content and to draw attention to the necessity of applying the paraphrase skill to gain meaning. To date, however, gloss notations have been constructed with "groups" of students in mind. To truly individualize this procedure, another method of glossing needs to be explored and refined: what is needed is a system which allows the reader to determine which gloss notation is needed for recovering comprehension and to activate it at that time. This way, each gloss offering is only presented when the student decides—through self-regulation—that the gloss is needed. Reader-activated gloss usage could serve well to insure the shift from teacher-guided to student-directed instruction. And a promising way to do this is to incorporate computer-assisted instruction into the glossing approach.

Computer-assisted instruction could provide the flexibility necessary to allow the reader to interact with the text and monitor specific gloss needs. The key word is "interact." Computer-assisted instruction is designed to provide instant service to the student (i.e., requesting and receiving a specific gloss notation). The computer has infinite patience so the student may proceed at a pace that is comfortable for that individual. Also, since the organization of the courseware is determined by the teacher, gloss paraphrase notations (in this case) can appropriately reflect that teacher's expectations and ideas. As more classrooms and homes become equipped with mini- and micro-computers for instruction, electronic media will replace much of the hardcopy versions of text. This shift from the printed page to the printed terminal screen presents unique opportunities for glossing instruction and research. My ultimate goal is to develop an independent programming system for computer-aided glossing that can be programmed by the classroom teacher who has little or no background in this area (Blohm, in progress). At the present, however, a prototype system has been developed that can be used with college students who have regular access to computer terminals to both guide text learning and to research treatment effects on that learning.

Guided by the purposes and rationale presented, the following questions were generated for this initial investigation. First, would the students who are provided the opportunity to activate gloss paraphrases while reading on the computer recall more than those who are not provided with the gloss paraphrase option? And second, would the number of gloss paraphrases activated influence the amount of recall?

METHOD

The design of the study was a 2×2 factorial with twenty college undergraduate Education majors randomly assigned to each of the four experimental groups. The first factor was the type of reading selection assigned to the subjects in each group which was either the topic of heat flow or electrical conductivity. The second factor was the type of treatment the subjects received which was either glossed paraphrases with the selection or no glosses. The sequencing of the treatment selections was counterbalanced by the computer such that two of the groups read the heatflow selection either with glossed paraphrases or without gloss paraphrases while the remaining two groups read the electrical conductivity selection either with gloss paraphrases or without. The treatment selections for the experiment were the "abstract" versions of two passages used in Royer and Cable (1975). The length of the selections in words was as follows: heatflow—780, electricity—672. The selections were then programmed in PL/1 computer-language as the text content of the computer-assisted courseware for use in the experiment. A gloss paraphrase/no-gloss paraphrase option was included for both selections in the "control" file of the courseware to automatically arrange the sequencing of the selection and treatment in the experiment. Separate "gloss-directions" and "no-gloss-directions" files were constructed respectively to guide and direct the subjects through the assigned selection and treatment. A record-keeping "dump-statistics" file was generated to store the subjects' activation of gloss paraphrases by page and sequence for subsequent analysis. (See Allen, Wyatt and Blohm, 1981, for a detailed description of the courseware employed.)

The subjects were run in groups ranging in size from one to five. All subjects in the experiment were instructed by their respective directions to read their assigned text for the purpose of recalling as much information given as possible. These directions explained that no screen pages could be reread after the subject signalled for a new page. The gloss paraphrase treatment directions for the two selections instructed those assigned subjects to activate *all* gloss paraphrases they feel would be necessary for helping them recall the text content; at the same time, however, they were encouraged to activate *only* those glosses they really need. Practice glosses were given in the directions for learning to signal the computer for the gloss. In contrast, the no-gloss treatment directions for the two selections simply explained to those subjects assigned that no more than eight lines of text would appear on the terminal-screen page at any one time. Following the separate introductory statements and procedures for the two treatments, all subjects were instructed in the courseware to read the treatment selection on the terminal using the assigned technique at their most comfortable reading rate. No time limits were imposed for reading. After they finished reading, the subjects were then asked to write down, in complete sentences, everything they could remember from the assigned selection. They were given as much time as they desired for completing this recall task. A graduate assistant who was naive to the purposes of the experiment scored each of the recall protocols against a list of "idea units" established in Royer and Cable (1975). The heat flow selection contained 67 idea units and the electrical conductivity selection contained 52. Twelve of the protocols (three from each group) were randomly selected and scored blindly by the author. The interscorer reliability coefficient for this sample was .97.

RESULTS AND DISCUSSION

Results of the study indicated that the group which received the reader-activated gloss paraphrase option on the computer for the heat flow article recalled more idea units on the free recall measure ($\bar{x} = 18.57$) than any of the other three groups ($\bar{x} = 16.15$, $\bar{x} = 12.21$, $\bar{x} = 7.15$). Both groups receiving the reader-activated gloss option recalled more idea units than did the groups not receiving the gloss option, regardless of selection read. Results of an analysis of variance indicated that the differences were significant between the mean idea units recalled by the two treatment groups ($p < .000$) for both selections ($p < .007$). No significant differences were found for the interaction between treatment and selection. Based on these results, students who were offered the option to activate gloss paraphrases while reading on the computer recalled more on the recall measure than did those who were not offered the reader-activated gloss option. Also found was that the students who read the heat flow selection recalled more than did those who read the electrical conductivity selection.

Results of Pearson Product Moment correlational analyses indicated that no significant correlation existed between the number of gloss paraphrases activated and the number of idea units recalled for either selection. The number of gloss paraphrases activated by students in the gloss option treatment while reading did not increase their performance on the reading recall measure. Therefore, the results indicate that the frequency of reader-activated gloss paraphrase usage did not influence the amount of recall.

An interpretation may be made of the significant treatment effect found for reader-activated gloss paraphrases when used while reading to facilitate recall. The gloss paraphrase, when activated by the student during reading on the computer, may have influenced the student's immediate comprehension by providing concrete restatements of the knowledge structures in the text which more closely matches that student's prior knowledge and experience. Since the paraphrases contain descriptions which explain the original technical knowledge in less technical terms, the student is able to relate the information to practical experience and previous learning more readily encoded. That no significant correlation existed might have been due to the comprehension-monitoring ability of the student which was promoted through the computer-aided treatment. Since the student was encouraged through the gloss paraphrase option to decide regularly while reading if the knowledge structures in the selection were being effectively encoded for later retrieval, the number of gloss paraphrases activated would depend on his or her prior experience with the knowledge contained in the selection. In addition, one activated gloss paraphrase may have had a facilitative transfer effect for understanding other segments of the original text. By gaining the intended meaning of one segment of the original text, the student may have been able to encode information following the segment glossed. Hence, the inclusion of the reader-activated gloss paraphrase option may have encouraged the student to not only monitor comprehension in terms of previous experience and learning but also in terms of actively relating earlier segments of text learning to later segments: an important metacognitive skill (Brown & DeLoache, 1978; and Yussen, Mathews, & Hiebert, in press).

An explanation for the significant difference found between the amount of recall by selection may be attributable to the number of idea units contained in the two treatment selections. The heat flow selection contained 67 idea units while the electrical conductivity selection contained 52; so a difference in recall by selection should be expected. That both selections were significantly better recalled by the reader-activated gloss option group substantiates the overall effectiveness of the treatment.

The results and interpretations of this investigation provide the foundation for further research and development of the computer-aided gloss option technique for ultimate implementation and experimentation in the classroom. Data from this experiment should first be reanalyzed for qualitative effects—depth of processing—as well as quantitative effects (McConkie, 1977). Then replications of the experiment should be conducted to increase the data-base for examining the effect of a particular activated gloss on the depth of processing exhibited in the recall of knowledge structures paraphrased to explain the effect of the activated gloss paraphrase. Incorporating additional features (e.g., backpaging, selection skimming) in the Compu-Gloss courseware as well as promoting more comprehension skills and strategies (e.g., seeing author's organization) in the reader-activated gloss notations may further improve the student's comprehension of text. Since the courseware is a flexible, independent program, these features can be readily incorporated for study. These findings and subsequent revisions in the courseware will then be downloaded for use on mini- and microcomputers for fieldtesting in classrooms at the various grade levels. The future for such hardware in the classroom is very encouraging. This study and related investigations are viewed as an attempt to capitalize on the use of the computer in the classroom for promoting reading skills and metacognitive strategies to facilitate students' ability to learn from text.

REFERENCES

Allen, R. J., Wyatt, S. Y., & Blohm, P. J. *Using the COMPU GLOSS System: An operator's manual* (Technical Report No. 105). Lafayette, Louisiana: University of Southwestern Louisiana, 1981.

Anderson, T. H. Study strategies and adjunct aids. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1980.

Baker, L., & Brown, A. L. Metacognitive skills of reading. In D. Pearson (Ed.), *Handbook of reading research*. New York: Longman, in press.

Blohm, P. J. *Development of the COMPU-GLOSS System for creating computer-aided gloss courseware*. Lafayette, Louisiana: University of Southwestern Louisiana, in progress.

Brown, A. L., & DeLoache, J. S. Skills, plans, and self-regulation. In R. S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

McConkie, G. W. *Studying retention from prose: A content structure approach*. Paper presented at the annual meeting of the American Educational Research Association, New York, 1977.

Otto, W., White, S., & Camperell, K. *Text comprehension research to classroom application: A progress report* (Theoretical Paper No. 87). Madison: Wisconsin Research and Development Center for Individualized Schooling, 1980.

Royer, J. M., & Cable, G. W. Facilitated learning in connected discourse. *Journal of Educational Psychology*, 1975, *67*, 116–123.

Royer, J. M., & Perkins, M. R. Facilitative transfer in prose learning over an extended time period. *Journal of Reading Behavior*, 1977, *9*, 185–188.

Yussen, S. R., Mathews, S. R., & Hiebert, E. Metacognitive aspects of reading. In W. Otto & S. White (Eds.), *Reading expository material*. New York: Academic Press, in press.

SUSAN M. BELMORE
PATRICIA MATTHEWS
CONNIE A. BRIDGE
SUSAN P. MOSKOW
SHEILA G. COHEN
University of Kentucky

EFFECTS OF INITIAL MENTION AND ORIENTING TASK ON READERS' COMPREHENSION OF MAIN IDEA

The purpose of this study was to investigate the effects of various types of orienting tasks upon subjects' comprehension of expository paragraphs, half of which had the main idea explicitly stated in an initial topic sentence and half of which had the main idea sentence omitted. Since comprehension can be viewed as interaction between reader and text, then readers' ability to comprehend and recall information from text can be affected by variables within the textual materials, within the readers themselves, and by factors which influence the nature of the interaction between reader and text.

The text variable in the present study involved the explicit statement of a main idea in the initial sentence in a paragraph. Previous studies have indicated that learners have difficulty comprehending and recalling paragraphs in which the initial mention convention is violated; that is, paragraphs in which the topic sentence is not presented at the beginning of the paragraph (Kieras, 1978, 1980). Several authors assert that the recall of the topic sentence has been due to the "levels effect"; that is, the information that is of high level importance in a propositional analysis of the content structure is better remembered than information that is of low level importance (Johnson, 1970; Kintsch, Kozminsky, Streby, McCoon, and Keenan, 1975; and Thorndyke, 1977). Meyer (1977) contends that since the most important information usually appears first in a passage, it will coincidentally be the initial information that is best remembered primarily because it is most important. Kieras' results provided some support for the "levels effect," but he also found that the same high importance proposition placed in another position in the paragraph was less influential than when it appeared in the initial position, indicating that factors other than levels of importance were also influencing readers' perception of the main idea.

Miller and Kintsch (1980) contend that the reason that comprehension is facilitated when a superordinate proposition is presented first is due to the "leading edge" strategy, which causes important propositions to be better remembered because they are processed more frequently. That is, the superordinate proposition occurs and is processed first after which ensuing propositions are processed by connecting them to the superordinate proposition on the basis of argument overlap. However, if a proposition lacks common arguments, it cannot be connected with the preceding superordinate. This leads to a search of long-term memory for a linking proposition which, if not found, requires a bridging inference to be made, thus resulting in greater processing demands and consequently a less comprehensible text. Thus, a paragraph which violates the initial mention convention would be one in which the superordinate proposition would not be presented first and the reader's attempts to connect the other propositions in the text to a superordinate would be disrupted. In the present study, subjects were asked to comprehend and recall paragraphs in half of which the initial mention convention was violated by the omission of the topic sentence.

In addition to the factors within the text, there are other factors which can influence the nature of the interaction between reader and text, such as the type of instructions given to the reader prior to encoding. Several studies have shown that when subjects are required to generate sentences about text paragraphs, they demonstrate superior comprehension (Doctorow, Wittrock, and Marks, 1978; Dee-Lucas and Divesta, 1980; and Taylor and Berkowitz, 1980). Furthermore, the influence of the orienting tasks have differed in some studies across reader ability and have shown differential effects on dependent measures which are sensitive to the type of information highlighted by the orienting task. For example, Doctorow, Wittrock, and Marks (1978) found that both good and poor sixth grade readers' comprehension was enhanced by generative instructions, but that poor readers benefitted more than good readers. They hypothesize that low-ability readers may not sponstaneously generate verbal elaborations for text as readily as do the high-ability readers, but when induced to do so by an orienting task, they perform more like high-ability readers. Based upon the findings of these previous studies, the present study was an attempt to look at the effect of type of orienting instructions on good and poor readers' comprehension and recall of the main idea of expository paragraphs, half of which had the main idea explicitly stated in an initial topic sentence and half of which had the main idea omitted.

Research Questions

The major research questions to be answered in this study were:

1) Would good and poor readers have more difficulty recalling the main idea of paragraphs in which the main idea sentence was omitted than paragraphs which contained an explicit statement of the main idea at the beginning of the paragraph?
2) Would type of orienting task (generation of a main idea statement, recognition of a main idea statement, or rating comprehensibility) affect good and poor readers' ability to recognize and recall the main idea of explicit and implicit paragraphs?
3) Would type of orienting task affect good and poor readers differently when reading explicit and implicit paragraphs?

It was expected that the main idea of a passage was more likely to be remembered if the learner had been forced to generate the main idea or if they had read an explicit statement of the main idea in an initial topic sentence. Additionally, it was hypothesized that the greater depth of processing induced by the generation task would compensate for the violation of the initial mention convention in paragraphs in which the main idea statement had been omitted. Thus, it was predicted that students who received instructions to generate a main idea statement would have superior recall and recognition of main ideas, especially on those paragraphs in which there was no main idea statement presented initially. In light of the findings of previous studies, it was also hypothesized that poor readers would especially benefit from instructions to generate and from the provision of an explicit main idea statement. These predictions were, of course, based upon the assumption that learners can be induced by the orienting generation tasks to perform the constructive activity and that they will be able to perform it appropriately and thus generate a logical main idea sentence.

METHOD

Subjects

A total of 54 undergraduates enrolled in introductory psychology at the University of Kentucky received course credit for their participation in this study. Eighteen subjects were assigned to each of the three conditions.

Materials

The stimulus passages were 16 paragraphs designed to illustrate the use of topic sentences in expository writing (Raygor, 1969). This type of material was used for several reasons. The paragraphs were all fairly short (approximately 4–5 sentences each), and the writing style was representative of college text materials. The content of the passages included a diversity of topics from history, science, and philosophy. Most importantly, the paragraphs were constructed in such a way that the first sentence of each was a "topic sentence" or a clear statement of the main idea of the passage.

An example of the stimulus passages follows. The topic sentence has been underlined.

> A star's color and brightness depend entirely on its temperature. A star like the sun contains many different elements, but these have nothing to do with the yellowish-white color of the sun. Rather, stars react to changes in temperature in the same way as metal objects do. A piece of iron at low temperatures glows a dull red. Heat it up and its color becomes cherry red, orange, yellow, and finally "white-hot." The tungsten filament in modern electric light bulbs shines with a bright white light because it is so hot; the carbon filaments in old-fashioned light bulbs were yellowish in color because they were much lower in temperature. Similarly, stars at different temperatures appear to be different colors.

Two versions of each stimulus paragraph were prepared: a *topicalized* version (in which the main idea sentence was included in initial position) and a *non-topicalized* version (in which the initial topic sentence was omitted). Each subject read the topicalized version of 8 of the paragraphs and the non-topicalized version of the remaining 8 paragraphs. Passage type was counterbalanced across subjects, so that an equal number of subjects saw each version of a given stimulus paragraph.

For each of the paragraphs, a forced choice item containing a pair of alternative statements was also constructed. One alternative was a correct statement of the main idea (a paraphrase of the topic sentence); the other alternative was a related but incorrect statement. For example, the above stimulus passage was tested by the following recognition pair:

a) The higher the temperature of a star, the brighter and more distinct its color. (correct)

b) The colors of both stars and man-made products depend on the elements of which they are composed. (incorrect)

A recognition test containing 24 items was also constructed for this experiment. The recognition items consisted of single statements to which subjects responded either "yes" or "no" to indicate whether they represented the main idea of one of the passages or not. There were three types of statements on the recognition test: Central (sentences which paraphrased the main idea of one of the passages), Noncentral (sentences which were related to one of the passages but were not main ideas), and Unrelated items (sentences which did not relate to any of the stimulus passages). There were eight sentences of each type representing randomly designated passages. Four items of each type referred to topicalized and four to non-topicalized passages.

Orienting tasks

Eighteen subjects were assigned to each of three orienting conditions: Generation, Recognition, and Rating. Students in the Generation condition were asked immediately after reading each paragraph to write down the main idea of each passage in their own words. Specifically, they were told to, "Write a sentence which summarizes the main idea or central points of the paragraphs. Write a sentence which you think would help someone understand what the paragraph is about even if they had not read it." In the Recognition condition, they were required to select the most appropriate main idea from a pair of similar statements. The Rating group simply rated how easy it was to understand the main idea of the passage on a scale from 1 (easy) to 7 (hard).

Procedure

Subjects were asked to read each passage carefully and encouraged to make a special attempt to understand the main idea. They were allowed to read through the paragraphs at their own pace, reading each paragraph once and completing the orienting task before going on to the next paragraph. All subjects in an experimental session were given the same orienting instructions. Except for the orienting tasks, the procedures and instructions were identical for all groups.

When all of the subjects in a given session had finished reading all 16 passages they were given an unexpected free recall test. They were asked to write down a sentence stating the main idea for each passage they could remember. They were cautioned to recall only the main idea for as many passages as possible, not the detailed contents of any particular passage. Five minutes were allowed for completion of free recall. Following recall, the Rate of Comprehension test (Van Wagenen, 1953) and the recognition test were given. On the recognition test, the subjects had to decide whether each statement expressed the main idea of one of the passages that they had read and circle "Yes" or "No" on the response sheet to indicate their decision. They also provided an estimate of their confidence in that response on a scale from 1 (very unsure) to 5 (very sure).

RESULTS AND DISCUSSION

Subjects in each orienting condition were divided into good and poor readers by performing a median split on their scores on the Van Wagenen Rate of Comprehension Scale. To ensure that the average reading levels of subjects randomly assigned to the different orienting conditions were equivalent, reading scores were subjected to a one-way ANOVA. The results indicated that the differences among the three groups with regard to reading ability were not significant. Thus, any group differences can be attributed to the nature of the orienting tasks and not to inherent ability differences between the groups.

Scores on each of the different orienting tasks were first analyzed separately. For the Rating group, the average comprehensibility rating across subjects (on a scale from 1—hard to 7—easy) was 3.31 for both the explicit passages and non-topicalized paragraphs. A two-way repeated measures ANOVA produced no significant main effects or interactions. Subjects at both levels of reading ability thus rated these passages as similar in comprehensibility regardless of whether the topic sentence was present or not.

In the Recognition condition, the probability of correct responses was greater for good readers ($\bar{X} = .91$) than poor readers ($\bar{X} = .78$), $F(1,16) = 5.02$, $p < .05$. Explicit passages ($\bar{X} = .88$) also produced more accurate performance than implicit passages ($\bar{X} = .81$), but this effect did not reach significance, $F(1,16) = 3.66$, $p < .07$. The interaction was not significant, indicating that the good readers in this condition were better at comprehending the main idea of both types of passages.

A scoring system was developed to score both the statements produced by Generation subjects immediately after reading and the statements in the free recalls of the subjects in all three groups. This system involved rating each main idea statement recalled on a scale from 0 to 4, based upon completeness and accuracy. The results of this analysis indicated that good readers ($\bar{X} = 3.13$) generally produced more complete and accurate main ideas than poor readers ($\bar{X} = 2.58$), $F(1,16) = 9.07$, $p < .01$. Neither the main effect for Passage Type nor the interaction was significant. To summarize the orienting task performance, good readers in both Generation and Recognition conditions tended to comprehend the passages more accurately than poor readers; however, both types of readers rated the paragraphs equally comprehensible in the Rating condition. There were no strong effects of topicalization on these orienting tasks except for a marginal superiority in the immediate recognition of the main idea for topicalized paragraphs.

For the recall data, average scores were computed for each subject by summing the quality ratings of the main ideas recalled from each type of paragraph, then dividing by the total number of passages of each type (8). The resulting data, which are summarized in Table 1, were subjected to a three-way (Orienting Task X Passage Type X Reading Ability) repeated measures ANOVA. This analysis revealed that recall was affected by the orienting tasks which subjects were assigned, $F(2,48) = 5.95$, $p < .01$. Pair-wise comparisons using Tukey's wholly significant difference (WSD) technique revealed that subjects in the Generation group performed significantly better than subjects in the Rating condition ($p < .05$), while the performance of the Recognition subjects was inter-

TABLE 1
Mean Quality of Ratings of Main Idea Statements in Free Recalls of Explicit and Implicit Paragraphs*

	Rating	Recognition	Generation
Poor Readers			
Explicit	0.406	0.933	1.377
Implicit	0.654	0.793	1.389
Good Readers			
Explicit	1.181	1.516	1.280
Implicit	1.131	1.508	1.654

*Recalls scored on scale of 0–4.

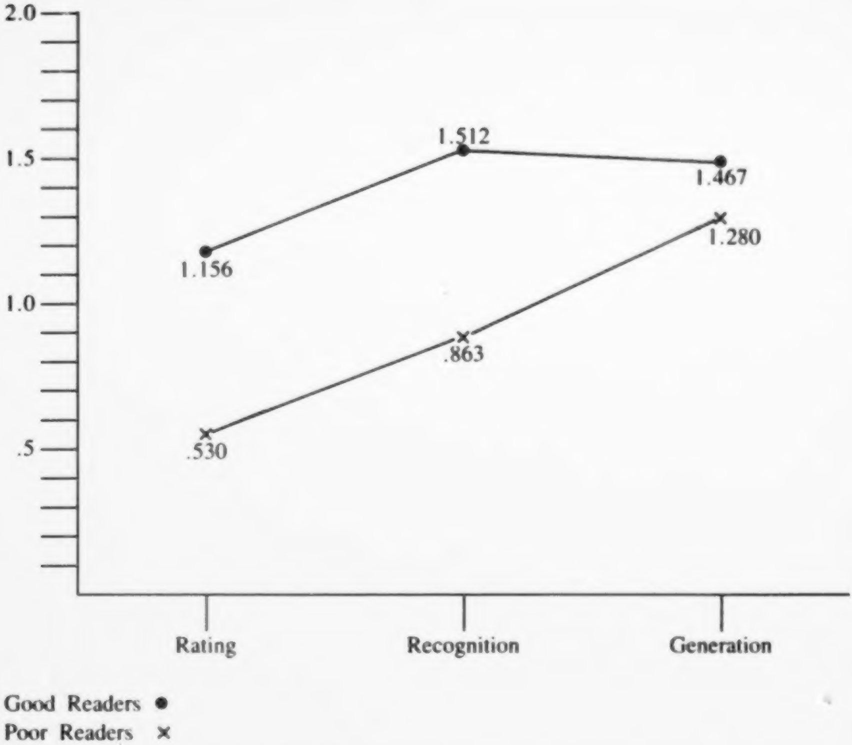


FIGURE 1. Mean Quality Ratings of Main Idea Statements in Free Recall.

mediate and did not differ significantly from the other two groups. The requirement to generate a main idea after reading a passage thus seemed to facilitate later recall of those ideas more effectively than rating comprehensibility but not more than selecting a main idea statement. Passage type was not significant either as a main effect or an interaction; recall was almost equivalent whether the main idea was explicit ($\bar{X}=1.11$) or implicit ($\bar{X}=1.19$). Good readers performed significantly better on the recall test ($\bar{X}=1.38$) than poor readers ($\bar{X}=0.92$), $F(1,48)=10.70$, $p<.01$. The marginal reading ability X orienting task interaction ($p=.18$) has been plotted in Figure 1. Examination of Figure 1 suggests that poorer readers were particularly helped by the Generation task. This observation was verified by *post hoc* tests which showed that for poor readers, performance in the Generation condition was significantly better than the Rating condition. Moreover, poor readers' recall was not significantly different from good readers when they were induced to integrate the information in the paragraph by generating a title sentence.

A slightly different pattern of results was observed with the recognition measure. Weighted recognition scores were computed by assigning a + to the confidence rating for a correct response and a − for an incorrect response, yielding scores which ranged from −5 to +5. Mean scores were calculated separately for the three item types (Central, Noncentral, and Unrelated) and analyzed in a repeated measures ANOVA. A preliminary analysis indicated no difference across conditions for the Unrelated items. Mean scores for the Central and Noncentral items were almost identical for topicalized ($\bar{X} = 1.56$) and nontopicalized ($\bar{X} = 1.58$) paragraphs. Good readers were somewhat more accurate than poor readers, but this difference did not approach significance. As expected, performance was much more accurate for correct Central items ($\bar{X} = 3.27$) than incorrect Noncentral items ($X = -0.12$), $F(1,48) = 75.95$, $p < .001$. Task instructions also had a significant effect on performance, $F(2,51) = 6.06$, $p < .01$. Follow-up analyses using Tukey's WSD procedure indicated that both the Recognition and Generation subjects were superior to subjects in the Rating condition ($p < .05$), but did not differ from each other. Recognition subjects were more accurate and/or more confident than other subjects in recognizing correct main ideas, although they showed no superiority in rejecting inaccurate or irrelevant statements. The reader is reminded that Recognition condition subjects were presented with an explicit statement of the main idea for each paragraph during the orienting task. Thus, it is hardly surprising that they should be better able to recognize a paraphrased version of the same idea later. The recall measure, on the other hand, showed that the Generation subjects did slightly better than the Recognition subjects in actually retrieving the main ideas of passages which they had been forced to construct earlier although this superiority was not significant.

CONCLUSIONS AND IMPLICATIONS

The data presented here indicate that the subjects in this study were able to overcome the disadvantage of reading paragraphs in which the initial mention convention was violated by the omission of the initial topic sentence. The activities induced by the orienting tasks appeared to compensate for the lack of an explicit main idea statement, thus supporting the hypothesis that reader-generated and author-provided themes can be remembered equally well when comprehension is induced by means of instructions which force the subjects to generate the appropriate main idea. This ability of readers to overcome the effects of texts which violate linguistic conventions is encouraging in view of a recent observation by Bransford (1981) that "the world is full of inconsiderate texts." An analysis of expository texts by Anderson, Armbruster, and Kantor (1980) supports the contention that students are frequently faced with the task of comprehending poorly written texts. Teachers may be able to help students deal with these materials through the use of instructions which require active involvement with the text.

These instructions may be especially helpful for poor readers who either do not possess or do not spontaneously use effective comprehension strategies. While the comprehension of good readers was better than that of poor readers in all three orienting task conditions and for both versions of the paragraphs, the performance of the poor readers did not significantly differ from that of the good readers when they were required to generate a main idea sentence for themselves. This indicates that the optimal type of instructions may differ for good and poor readers. The good readers appear to comprehend well regardless of orienting task and probably apply their own strategies spontaneously; while the poor readers may need to perform more active tasks to comprehend textual material. When they do comprehend, they no longer seem to be at a disadvantage in memory for the main ideas of that they read.

REFERENCES

Anderson, T. H., Armbruster, B. B., & Kantor, R. N. *How clearly written are children's textbooks? Or, of bladderworts and alfa* (Reading Education Report No. 16). Champaign, Illinois: University of Illinois, August 1980.

Bransford, J. *Reading and learning*. Paper presented at the meeting of the College Reading Association, Louisville, October 1981.

Dee-Lucas, D., & Di Vesta, J. J. Learner-generated organizational aids: effects on learning from text. *Journal of Educational Psychology*, 1980, *72*, 304–311.

Doctorow, M., Wittrock, M. C., & Marks, C. Generative processes in reading comprehension. *Journal of Educational Psychology*, 1978, *70*, 109–118.

Johnson, R. E. Recall of prose as a function of the structural importance of the linguistic units. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 12–20.

Kieras, D. E. Good and bad structure in simple paragraphs: effects on apparent theme, reading time, and recall. *Journal of Verbal Learning and Verbal Behavior*, 1978, *17*, 13–28.

Kieras, D. E. Problems of reference in text comprehension. In P. Carpenter & M. Just (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Erlbaum, 1980.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan J. J. M. Comprehension and recall of test as a function of content variable. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 196–214.

Meyer, B. J. F. What is remembered from prose: A function of passage structure. In R. O. Freedle (Ed.), *Discourse production and comprehension: Advances in research and theory* (Vol. 1). Norwood, N.J.: Ablex, 1977.

Miller, J. R., & Kintsch, W. Readability and recall of short prose passages: a theoretical analysis. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, *6*, 335–354.

Raygor, A. L. *Reading for the main idea*. New York: McGraw-Hill, 1969.

Taylor, B., & Berkowitz, S. Facilitating children's comprehension of content material. In M. L. Kamil and A. J. Moe (Eds.), *Perspectives in reading research and instruction*. Clemson, S.C.: National Reading Conference, 1980.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, *9*, 77–110.

Van Wagenen, J. J. Rate of comprehension scale, form A. Van Wagenen, M. J. Minneapolis, Minnesota, 1953.

EUGENE H. CRAMER
University of Illinois

READING COMPREHENSION AND ATTITUDE OF HIGH AND LOW IMAGERS: A COMPARISON

Two continuing problems that have been the subject of considerable research are the development of reading comprehension ability and the cultivation of positive attitude toward reading. In recent years, a number of studies have found a positive relationship between reading comprehension and mental imagery (Anderson, 1972; Dillingofski, 1981; Koskinen & Gambrell, 1980; Lesgold, Curtis, DeGood, Golinkoff, McCormick & Shimron, 1974; Levin, 1973). Much less work has been reported in the area of mental imagery and reading attitude (Cramer, 1980; Irwin, 1979). What is of interest is that researchers in both areas of reading comprehension and reading attitude have been working with the concept of mental imagery. This study attempted to address both reading comprehension and reading attitude in their possible relationships with mental imagery.

The concept of mental imagery is a very old idea which is currently undergoing a revival of interest. Mental imagery has been defined by Litt (1973) as revived sensory experience in the absence of the original sensory stimulation. Any study that undertakes to investigate the concept of mental imagery must deal with an important obstacle: the reliability of introspective reports of imaging. The work of Marshall (1968) and the earlier work of Golla, Hutton, and Walter (1943) in the measurement of alpha rhythms through electro-encephalographic recording during periods of self-reported imaging has done much to establish the validity of the mental imagery concept.

The purpose of this study was to investigate relationships among the three major concepts, reading comprehension, attitude toward reading, and mental imagery. The principal research question asked in this study was, "If a reader self-reports a high degree of mental imaging, will that reader also score well on comprehension tests of prose passages and report a high attitude toward reading?" In short, if it were established that mental imagery is related to both reading comprehension and reading attitude, then many questions of major importance to the teaching of reading could be addressed. For example, is the ability to form mental images susceptible to instruction during reading? If so, what are effective and efficient methods of developing mental imaging during reading? What are the effects of mental imaging on long-term retention of material learned through reading? What are long-term effects of imaging ability on positive attitudes toward reading? Are there optimum ages for developing reading comprehension skills or for nurturing positive attitudes through the use of mental imagery? All of these and many other related questions would, however, be pointless without a demonstrable correlation among the three major concepts of reading comprehension, reading attitude, and mental imagery.

METHOD

Subjects

The subjects of the study were 30 eleventh-grade students, selected randomly from language arts classes in two high schools (one urban, one suburban) in the Chicago area. In each school, the 15 subjects were characterized by local ad-

43

ministrators and reading specialists as at or slightly below average achievement levels. No other demographic data were collected for the subjects since it was believed that effects of such variables as sex, race, IQ, and age would be controlled through the randomization process, and since it was also believed that the concepts of mental imagery, reading comprehension, and reading attitude ought to be powerful enough to emerge from others. It would be fair to say that among the 30 subjects were representatives of several ethnic groups and an almost equal balance of males and females.

Materials

Estes Reading Attitude Scale. Developed by Estes (1971), this attitude measure was validated in a study by Dulin and Chester (1974). The Estes Reading Attitude Scale was administered to all subjects by their normal classroom teachers three days before the other measures were given to the subjects. This procedure was followed so that (a) subjects' attitudes could be compared with those of their regular classmates and (b) subjects' attitudes would not be influenced by other measures used in this study.

Prose comprehension tasks. Two prose passages of approximately 300 words were selected for the study. The first passage, which was about a tiger hunt, was estimated to be of a high image-evoking nature; the second passage, which was a discussion of the theory of hypnosis, was estimated to be of low image-evoking quality. The designations of high image-evoking or low image-evoking quality were verified through an inter-judge reliability procedure conducted with graduate students in education who were in unanimous agreement with the original designations. Both passages were also judged to be within the readability limits of average ninth grade students.

Comprehension was assessed in two ways. First, a five-item short-answer test was constructed for the first paragraph of each of the two passages; i.e., a five-item test for the high-image-evoking passage, and a five-item test for the low-image-evoking passage. Each short-answer test contained three literal recall questions and two inferential questions. Question type and passage dependency of all questions was verified through an inter-judge reliability procedure. Short-answer tests were chosen over other traditional comprehension measures, such as multiple-choice or true-false questions, in order to eliminate or at least control the "guessing" factor. Subjects' answers to the short-answer items were marked as either correct or incorrect independently by the investigator and three graduate students in reading education. Initial scoring achieved near-perfect agreement among the judges, and the few minor discrepancies were resolved through group decision.

In addition, for each of the two passages, low and high image-evoking, a modified cloze test was constructed from the second paragraph of the passage. This was done as a further measure of subjects' reading comprehension beyond the short-answer tests to either verify the short-answer test results or determine whether the results of cloze passages or short-answer tests might somehow differ for high and low imagers. Since subjects had just finished a 300-word passage for the short-answer test, it could be assumed that they were well-attuned to the content of the passage and the author's style of writing; thus it was decided to reduce the length of cloze passages to twenty blanks in an effort to control for subject fatigue. Scoring of the cloze tests was on the basis of exact replacement of missing words in an every-fifth-word deletion pattern for a total of twenty blanks in each passage.

Visual-recall task. In the visual-recall tasks, unlike the prose comprehension tasks, subjects were specifically instructed to form mental images about what they were reading in order to perform a simple diagramming task without referring to the passage. Three visual-recall tasks were adapted from workbook-type exercises. Task one involved three items which were part of directions from a filling station to a restaurant. Task two involved seven items involving details of geography of an island. Task three involved ten items dealing with the physical arrangement of objects in a country garden. Thus, there were twenty items over the three visual-recall tasks for subjects to draw in simple diagrams. Each item remembered and placed correctly in the diagram was scored as one point.

Questionnaire on reading activities. Subjects completed a brief questionnaire immediately upon completion of the prose comprehension tasks and prior to attempting the visual-recall tasks. In the questionnaire, subjects stated whether or not they had experienced mental imaging during the prose comprehension tasks, how much of the time imaging occurred, if indeed it had, which passage seemed easier, which passage they had enjoyed more, and whether or not they like to read.

Procedure

Three days after subjects had been administered the Estes Reading Attitude Scale as a part of their normal classroom activities, they were given the prose comprehension tasks, the questionnaire on reading activities, and the visual-recall tasks. On the basis of subjects' responses to the imaging questionnaire, three groups of data were derived: high imagers were those subjects who reported using imagery 75–100% of the time they were reading the prose comprehension passages; moderate imagers, 50%, and low imagers, 0–25%.

RESULTS AND DISCUSSIONS

Mean scores and standard deviations were calculated for the three imagery groups (high, moderate, and low imagers) on the attitude measure and the comprehension and visual recall tasks. Results, as shown in Table 1, indicate that in every case high imagers performed at the highest level, moderate imagers at a middle level, and low imagers at the lowest level on both of the prose comprehension tasks and on the visual-recall tasks. Although only one comparison of mean scores reach statistical significance (short answer task for low image-evoking passage), a clear trend is apparent, stepwise in the anticipated direction.

Perhaps the most interesting finding in this study was in the area of reading attitude. A simple ANOVA was calculated for the three imagery groups on the attitude variable with the resultant $F(2,27)=3.56$, $p<.041$. However, of special interest was the lowest attitude mean score of the moderate imagery group. Further, the standard deviation of the attitude score of the moderate imagery group was about twice as great as that of either of the other two groups.

Therefore, the main question addressed in this study is answered with a tentative affirmative, since the results generally support the idea that mental imagery is related not only with reading comprehension but also with reading attitude. Although mean scores of high, moderate, and low imagery-reporting subjects

TABLE 1
Means and Standard Deviations on an Attitude Scale and Five Comprehension Tasks by High, Middle, and Low Imagery Groups

		*High Imagers** (75–100%)	Mid Imagers* (50%)	Low Imagers* (0–25%)
ATTITUDE SCALE (range = 20–100)	M	75.7	59.7	68.4
	SD	8.6	18.3	8.4
HIGH IMAGE PASSAGE (short answer task total possible = 5)	M	3.1	2.8	2.0
	SD	1.1	1.8	1.7
HIGH IMAGE PASSAGE (cloze task, total possible = 20)	M	9.2	7.3	6.1
	SD	1.3	4.1	3.6
LOW IMAGE PASSAGE (short-answer task total possible = 5)	M	3.2	2.2	1.5
	SD	1.6	1.5	1.3
LOW IMAGE PASSAGE (cloze task, total possible = 20)	M	9.0	7.8	6.4
	SD	2.1	4.8	4.2
VISUAL RECALL TASK (total possible = 20)	M	13.2	11.2	10.3
	SD	4.2	7.1	6.0

*High Imagers (n = 9)
Mid Imagers (n = 10)
Low Imagers (n = 11)

on comprehension tasks involving both high and low image-evoking prose passages were not significantly different, they did show a stepwise trend in the anticipated direction overall. Further, high imagers outperformed moderate and low imagers on visual recall tasks, also in a stepwise trend.

The fact that the moderate imagery group had the lowest mean score on reading attitude is a rather surprising finding which obviously requires further and more careful study. One speculation about this result is that moderate imagers seem much less certain about their reading attitudes than do either high or low imagers who have apparently made more definite decisions; this speculation is supported by the much greater standard deviation on attitude of the moderate imager group. However, the many complex relationships of mental imagery with reading comprehension and attitude could not be addressed in this study but may be teased apart in subsequent ones. Possible explanations for this result may be found when certain questions posed earlier in this paper are investigated: 1) is the ability to form mental images during reading susceptible to instruction? 2) are there optimum ages for increasing reading attitude with imagery strategies? and 3) what are the long-term effects of imagery strategies on reading attitude?

These and many other attendant questions would seem to require attention on the basis of this study and other previous ones referred to above. A number of limitations to this investigation should be noted. First, trends were noted, not significant differences; second, subjects' reading abilities were not controlled and, therefore, differences among high, average, and low ability readers relative to their mental imagery and attitude scores were not investigated; and finally, no causal influences among mental imagery, reading comprehension, and reading have been noted, nor should they be inferred from this relational study.

However, the results of this study seem sufficiently positive to warrant further investigations of mental imagery as it relates to reading comprehension and to reading attitude.

REFERENCES

Anderson, R. C. & Kulhavy, R. W. Imagery and prose learning. *Journal of Educational Psychology*, 1972, *63*, 242–243.

Cramer, E. H. Mental imagery, reading attitude, and comprehension. *Reading Improvement*, 1980, *17*, 135–139.

Dillingofski, M. S. *The effects of imposed and induced visual imagery strategies on ninth grade difference-poor readers' literal comprehension of concrete and abstract prose*. Paper presented at the twenty-sixth annual convention of the International Reading Association, New Orleans, April 1981.

Dulin, K. L. & Chester, R. D. A validation study of the Estes Attitude Scale. *Journal of Reading*, 1974, *18*, 56–59.

Estes, T. H. Scale to measure attitudes toward reading. *Journal of Reading*, 1971, *15*, 135–138.

Golla, F. L., Hutton, E. L. & Walters, W. G. The objective study of mental imagery I—physiological concomitants. *Journal of Mental Science*, 1943, *89*, 216–223.

Irwin, J. W. College readers' mental imagery, reading ability, and attitude toward reading. *Reading Improvement*, 1979, *16*, 124–129.

Koskinen, P. & Gambrell, L. The effects of an induced visual imagery strategy upon the reading comprehension of below average readers. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives in reading research and instruction*. Twenty-ninth Yearbook of the National Reading Conference. Washington, D.C.: The National Reading Conference, 1980.

Lesgold, A. M., Curtis, M. E., DeGood, H., Golinkoff, R. M., McCormick, C, & Shimron, J. *The role of mental imagery in text comprehension: Preliminary studies*. ERIC Document Reproduction Service No. ED094317, 1974.

Levin, J. R. Inducing comprehension in poor readers: a test of a recent model. *Journal of Educational Psychology*, 1973, *65*, 19–24.

Litt, L. Mental imagery and reading. *Reading: A journal for the study and improvement of reading and related skills*, 1973, *7*, 27–33.

Marshall, W. A. *Development of the brain*. Edinburgh: Oliver & Boyd, 1968.

JAMES W. CUNNINGHAM
University of North Carolina-Chapel Hill

GENERATING INTERACTIONS BETWEEN SCHEMATA AND TEXT

Instructional strategies, like jokes and recipes, come from somewhere, but it is rarely clear where that "somewhere" is. This paper reports an experiment in which an original teaching strategy was tested against a placebo. The strategy, the GIST procedure, was developed by the author to improve students' abilities to comprehend the gist of paragraphs.

BACKGROUND

Day (1980; Brown, Campione, & Day, 1981) increased learning from texts (as measured by ability to summarize them) for college students by teaching an explicit set of summarization rules based on the macro-rules of van Dijk (1977) and van Dijk and Kintsch (1978). (Brown (1980) had argued for making rules explicit to learners for metacognitive reasons.) Doctorow, Marks, and Wittrock (1978) and Taylor and Berkowitz (1980) both found that having students write a one-sentence summary following each paragraph they read dramatically increased learning from texts. Mason (1973) reported having led individual remedial readers to get the main idea of a paragraph by first dealing with just one sentence, then with two, etc.

Kintsch (1977) argued that while readers use cues from texts in the form of propositions and the order in which they are presented, extensive schema use is required for the reader to organize these propositions into the macro-structure of the text. Dooling and Christiaansen (1977) and Anderson (1978) have explained how what the reader already knows about the topic of a text influences what is understood and remembered from that text.

This investigator combined Mason's (1973) technique with Doctorow, Marks, and Wittrock's (1978), added guidelines to make it a teacher-directed small-group strategy, and provided a prescription for fading from group sentence-by-sentence gist production to individual whole-paragraph gist production. The strategy which resulted is the GIST procedure.

THE GIST* PROCEDURE (PARAGRAPH VERSION)

Step 1

Select paragraphs (each of which appears to have a gist) of three to five sentences in length at the appropriate level of difficulty for the students with whom the procedure is to be used.

Step 2

Present the first sentence of one paragraph to the students (on a chalkboard, an overhead transparency, or a chart). On a chalkboard, put 15 blanks. Tell the students to read the sentence so that they can retell it in their own words in one statement of 15 or fewer words.

**Generating Interactions between Schemata and Text.*

Step 3

When the students have finished reading the sentence, cover it. Say "I asked you to read the sentence so that you could retell it in a statement of 15 or fewer words. Who can start?" Writing one word per blank, allow the students as a group to dictate and edit the statement until it is complete. At any time that the students wish to see the original sentence, show it to them but they must dictate and edit from memory. If the students do give an exact duplicate of the sentence in the text, do not be disturbed. This step is complete when the students feel that their statement is an accurate retelling of the sentence in the text. The teacher enforces the rules of a single statement with 15 or fewer words but the teacher does not evaluate the content of the statement. The students are required to compare their final statement with the text. If they are satisfied, the teacher should be.

Step 4

Uncover the second sentence of the paragraph (leaving the first sentence uncovered). Erase the chalkboard and put 15 blanks. Tell the students to read the two sentences so they can retell them in their own words in one statement of 15 or fewer words.

Step 5

When the students have finished reading the sentences, cover them. Say "I asked you to read the sentences so that you could retell them in a statement of 15 or fewer words. Who can start?" Use the same means for leading them to construct their statement as were used in step 3.

Step 6

Continue this procedure, adding one sentence at a time, until the students have produced a single statement of 15 or fewer words which they feel captures the essence of the paragraph. Use this strategy with as many different paragraphs as necessary until you feel the students are becoming adept at efficiently producing gist statements for paragraphs as a group. At this point, uncover the entire paragraph at the beginning of the lesson and have the group construct a gist statement for the whole paragraph without building up to it sentence-by-sentence. When the students as a group become adept at producing gist statements for paragraphs presented as wholes, then have the students individually produce gist statements for paragraphs presented as wholes.

METHOD

The GIST procedure was tested against a placebo which afforded subjects equal experience with the training paragraphs and with composing-writing but which focused their attention on the meaning of individual words in the paragraphs.

Materials

The first 110 paragraphs from a piece of reading instructional material were selected (Boning, 1970). The exercises accompanying these paragraphs were deleted, leaving only the paragraphs themselves. The paragraphs were all by the same author, were all at the same level of reported difficulty (hard-third-grade level), and had all been produced to meet the same curricular goal (to teach students how to draw conclusions). Graduate students enrolled in their first graduate reading education course were asked to read some of these paragraphs silently and immediately write a one-sentence gist statement of 15 or fewer words for each one they read. These students were trained in advance to produce gist statements which were defined as "text-based summaries." Each of the 110 paragraphs was read by four graduate students.

The investigator and another reading researcher read the four gist statements for each of the 110 paragraphs and selected the paragraphs where both felt that at least three of the four gist statements were in substantial agreement. Sixty-seven paragraphs remained at the end of this procedure.

The investigator, using his experience with text analysis procedures (Frederiksen, 1975) produced a consensus gist statement for each of the 67 paragraphs remaining. During this process, the investigator included the modal word concepts and case relations of the four gist statements using the modal word choices and syntactic patterns of the four gist statements. The consensus gist statements were checked by another reading researcher. Seven more paragraphs were deleted during this process, leaving a corpus of 60 paragraphs and their consensus gist statements. The 40 paragraphs actually needed for this study were randomly selected from this corpus and were, in turn, randomly assigned to four sets of 10 paragraphs each.

All of the discrete words from the first set of 10 paragraphs were catalogued and 30 of these were randomly selected. These 30 words were written in manuscript, one to a card, on unlined white 3×5 index cards.

Subjects

All 121 fourth graders present on the first day of testing at an elementary school in the Southeast were screened. The children were given a gist recognition test over the 10 paragraphs from the first paragraph set. Booklets were prepared and administered so that each child would read the first paragraph, roll it over, and select from five choices which statement best told what the whole paragraph was about. The correct answer was the consensus gist statement for each paragraph; the incorrect answers were constructed by an algorithm to differ from the correct answer in systematic ways. Once a student had read a paragraph and rolled it over, s/he could not look back. Students were given ample time to complete the test.

On subsequent days, the children were tested for accurate and fast word identification of the 30 words selected from the first paragraph set. Each word was flashed for approximately .4 of a second and then the student was given between 1 and 2 seconds to report the identification of that word. The student's first response was taken even if s/he later corrected it. Twenty-eight students scored 86.7% (26 of 30) or higher on the test of accurate and fast decoding for the words in the paragraphs, but scored between 20% and 80% on the gist recognition test for the paragraphs. Twenty percent recognition of gist statements was considered too low because it was the chance score; eighty percent was considered too high because of lack of "room" to show growth. These 28 subjects were randomly assigned to two groups, experimental and placebo.

Procedures

Treatments. The second set of 10 paragraphs was used in the two treatments and the investigator was the teacher of both groups. Students were taught in groups of 14 in a small room in the school. The experimental group was taught by the GIST procedure to produce gist statements for paragraphs. The placebo group was taught by a variety of strategies which focused their attention to the word level of the paragraphs and which had them do at least as much writing as the experimental group. The placebo group had access to the same paragraphs as the experimental group for as long a period of time. Both groups had nine, 25-minute lessons spread across three weeks. All lessons took place in the morning.

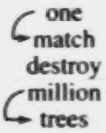
Dependent Measures. The third set of 10 paragraphs was used to construct a gist recognition measure similar to the pretest/screening gist recognition measure. The fourth set of 10 paragraphs was used to construct a gist composing-writing measure in which students were asked to read a paragraph, roll it over, and write one sentence telling what the whole paragraph was about. None of the paragraphs in either of the dependent measures had been seen before by the subjects. Students were given ample time to complete both tests.

Scoring. The gist recognition measure was scored as the pretest/screening gist recognition measure had been: the number of consensus gist statements selected from distractors across the 10 paragraphs.

The gist composing-writing measure was scored by a simplified text analysis procedure. For each of the 10 paragraphs, the consensus gist statement was turned into a set of scoring points by writing down each content word of the consensus gist statement in order. Content words linked to others by attributive or manner relations in the original text were linked in the scoring points by directional arrows so that a student could not receive credit for an attribute or manner unless it was coupled with the object, action, etc. to which it belonged. For example, the consensus gist statement

One match can destroy a million trees.

was rendered into the following scoring points:



The scoring points for the ten paragraphs were used by the investigator and another reading researcher to independently score all composed gists for all subjects. Subject protocols were scored for content only, not form. One point was given for each scoring point represented in a protocol. The interrater reliability for the 68 scoring points across the 28 subjects was .953 before consultation and .999 after consultation.

RESULTS

The gist recognition dependent measure was analyzed with a 2×2 analysis of variance with a repeated measure as the second factor. The first factor was experimental versus placebo; the second factor was pretest versus posttest. The main effect for group was not significant ($F=0.06$; $p<.82$); the main effect for

TABLE 1
Means (and Standard Deviations) of Gist Recognition Measures

Group	Time	
	Pretest	Posttest
Experimental	4.9(1.3)	6.3(1.9)
Placebo	5.1(1.5)	5.8(2.2)

TABLE 2
Means (and Standard Deviations) of Gist Composing-Writing Measure

Group	Raw Scores	Adjusted by Covariate
Experimental	33.7(6.1)	33.9
Placebo	27.9(8.3)	27.8

the repeated measure was significant ($F=9.53$; $p<.01$); the interaction effect (the test of interest) was not significant ($F=1.21$; $p<.29$). See Table 1 for means and standard deviations.

The gist composing-writing measure was analyzed with a *t*-test and with a 1×2 analysis of covariance using the pretest/screening gist recognition measure as the covariate. The *t*-test was significant ($t=2.11$; $p<.05$); the main effect for the ANCOVA was significant ($F=5.38$; $p<.03$). See Table 2 for means and standard deviations. (For the ANCOVA, the correlation between covariate and dependent variable was .33; the F-test of non-parallelism of regression was not significant.)

DISCUSSION

The subjects of this study were fourth graders from an elementary school in the Southeast who were relatively stronger in accurate and fast word identification than they were in paragraph gist recognition tested in a text-unavailable format. Results may not generalize to other populations of readers or to other types or instances of passages or to other tests. Within these limitations, however, certain cautious conclusions may be drawn.

The first conclusion is that the GIST procedure can lead a group of students like those used in this study to better recognize gist statements from distractors but that the improvement is not greater than would be obtained by providing students with exposure to paragraphs in the presence of composing-writing and strategies which focused their attention on individual word meanings in the paragraphs. Either the GIST procedure is only as effective at the placebo treatment for improving gist recognition or the GIST procedure would have to be used with students for a longer period of time.

The second conclusion is that the GIST procedure can, in a relatively short period of time, lead a group of students to compose more accurate gist statements for paragraphs than they would be able to after equal exposure to the same training paragraphs and equal time spent composing-writing. The effect size for the gist composing-writing variable on the experimental versus placebo factor was .7, indicating a large effect. This effect is especially important given that it was manifest on novel paragraphs not seen before.

Finally, students' improved abilities to compose gist statements for paragraphs was brought about by a method which, though systematic, did not teach explicit rules. Thus, though it is interpretable in light of current theory in psychology and linguistics, it is not dependent on that theory. A good teaching method ought to be interpretable in light of whatever theory is in vogue!

REFERENCES

Anderson, R. C. Schema-directed processes in language comprehension. In A. M. Lesgold, J. W. Pellegrino, S. D. Fakkema, & R. Glaser (Eds.), *Cognitive psychology and instruction*. New York: Plenum Press, 1978.

Brown, A. L. Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Brown A. L., Campione, J. C., & Day J. D. Learning to learn: On training students to learn from texts. *Educational Researcher*, 1981, *10* (2), 14–21.

Day, J. D. *Teaching summarization skills: A comparison of training methods*. Unpublished doctoral dissertation, University of Illinois, 1980.

Doctorow, M., Marks, C., & Wittrock, M. Generative processes in reading comprehension. *Journal of Educational Psychology*, 1978, *70*, 109–118.

Dooling, D. J., & Christiaansen, R. E. Levels of encoding and retention of prose. In G. H. Bower (Ed.), *The psychology of learning and motivation*. New York: Academic Press, 1977.

Frederiksen, C. H. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology*, 1975, *7*, 371–458.

Kintsch, W. On comprehending stories. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Mason, G. E. Personal communication. Athens, GA: 1973.

Taylor, B., & Berkowitz, S. Facilitating children's comprehension of content material. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives on reading research and instruction* (Twenty-ninth Yearbook of the National Reading Conference). Washington, DC: National Reading Conference, 1980.

van Dijk, T. A. Semantic macro-structures and knowledge frames in discourse comprehension. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

van Dijk, T. A., & Kintsch, W. Cognitive psychology and discourse: Recalling and summarizing stories. In W. U. Dressler (Ed.), *Current trends in textlinguistics*. New York: Walter de Gruyter, 1978.

MATERIALS

Boning, R. A. *Specific skill series: Drawing conclusions, book C*. Baldwin, NY: Barnell Loft, Ltd., 1970.

THOMAS H. ESTES
University of Virginia

THE SKILLS OF STUDYING

We know little of what people do when they study—when they are completing a homework assignment or when they are preparing for a test. As a way of gathering information to satisfy our curiosities about what people might at least *say* they do under these conditions, a thirty item inventory of study habits was administered to 418 college students, 48 tenth graders, and 76 seventh graders. All of these students were of average to above average reading ability. Many of the items were adopted from the 1941 edition of the *Study Habits Inventory* by C. Gilbert Wrenn.

There were three major questions of interest motivating this study:

1) Are there identifiable study habits factors which would describe the constituents of study practices? An ancillary question here is whether these empirically derived factors might be similar to those Tom Anderson and Bonnie Armbruster (1980) have called state variables and processing variables, or those Don Dansereau and his colleagues (1978) have called primary strategies and support strategies.

2) Are reported habits of study different under conditions where students are doing homework and under conditions where they are studying for a test?

3) In what ways might the reported study habits of seventh and tenth grade and college level students differ?

The questionnaire (reproduced in its post-experimentally revised form at the end of this article) asked students to respond "rarely or never," "sometimes," or "often or always" to a series of statements of study habits, first with reference to reading a homework assignment and second with reference to studying for a test. (Over 95% of the students questioned said that they at least sometimes read an assignment once when it is first given as homework and then review it one or more times in preparation for a test. One of our purposes in asking this preliminary question was to establish a proper mental set for our questionnaire, as well as to justify the distinction we were making.) Our initial analysis gave us proportions of students responding in each response category with respect to each condition. To achieve a more parsimonious view of the data, a classical factor analysis was conducted, repeating the procedure for odd numbered and even numbered subjects to assess factor replicability. Each item was "scored" by assigning a value of 1 to "rarely or never," 2 to "sometimes," and 3 to "often or always."

Looking at the studying for homework condition, a three factor solution seemed appropriate. On the odd/even split, two of these factors replicated at the 0.97 level, the remaining one at the 0.80 level, suggesting that in another sample the third factor might not reappear. It seemed, however, unduly strict if not rash to disregard the factor.

Looking at the studying for a test condition, a two factor solution seemed most defensible. One of the studying for homework factors seems to replicate in the studying for a test condition and one new factor seems to emerge in this condition. One of the studying for a test factors had an odd/even replication of 0.97, the other had a replication of 0.95.

An Inventory of Study Habits

	When reading a homework assignment			When studying for a test		
	Rarely or never	Sometimes	Often	Rarely or never	Sometimes	Often
1. I skim the reading to get a sense of the main ideas immediately before I read more carefully. [a]						
2. After skimming, I ask myself questions about the information in the text before reading more carefully. [a]						
3. I try to judge how the writer's ideas make sense. [c]						
4. While I read, I try to answer my own questions and the questions at the end of the chapter. [a,d]						
5. I try to relate the writer's ideas to things I learned in other places. [c]						
6. I try to memorize word-for-word most of the information. [a]						
7. I try to understand the text well enough to put it in my own words without losing the author's meaning. [c,d]						
8. I underline or set off in some way some parts of what I'm reading. [c,d]						
9. I take notes on what I'm reading. [a,d]						
10. I test myself by answering questions I've made up. [a,d]						
11. I test myself by answering end of chapter questions. [a,d]						

An Inventory of Study Habits

	When reading a homework assignment			When studying for a test		
	Rarely or never	Sometimes	Often	Rarely or never	Sometimes	Often
12. I try to put the main points of the passage in my own words. [c,d]						
13. I find it hard to keep my mind on what I am studying; I don't know what I have been reading about when I get through. [b]						
14. I go over the passage page by page and try to memorize the important points and details on each page. [a]						
15. I have a tendency to "day-dream" when trying to study. [b]						
16. I study with friends and we ask each other questions as well as discuss the material. [a]						
17. It takes me some time to get "warmed up" to the task; I am likely to waste time. [b]						
18. My study periods are too short for me to get "warmed up" and concentrate. [b]						
19. After having already read the passage carefully one time, I reread the complete passage carefully at least one more time. [a,d]						
20. I reread parts that I understand or set off in some way. [a,d]						
21. My periods of study are interrupted by outside interference such as telephone calls, visitors, and noises. [b]						

An Inventory of Study Habits

	When reading a homework assignment			When studying for a test		
	Rarely or never	Sometimes	Often	Rarely or never	Sometimes	Often
22: I reread my notes on the passage. [a,d]						
23. I find it hard to force myself to finish work by a certain time, under pressure. [b]						
24. I have trouble picking out the important points in material read or studied. [b]						
25. I have to wait for the mood to strike me or for an inspiration before starting a task; I am likely to waste time. [b]						
26. I read the questions at the end of the chapter before I read the assignments. [d]						

Key to factors loading:

a = loads on Industry/Compulsivity
b = loads on Distractability
c = loads on Initiative
d = loads on Assimilation

Consider the three factors which emerged for the studying for homework condition:

One factor had seven items clearly loading on it (in the range of 0.66 to 0.44) all with the tone and flavor of: "It takes me some time to get 'warmed up' to the task; I am likely to waste time"; "I have a tendency to daydream when trying to study"; "I find it hard to keep my mind on what I am studying; I don't know what I have been reading about when I get through." Clearly this factor is akin to Armbruster and Anderson's "state" variables and to Dansereau's "support" strategies, especially "concentration." The name chosen for the factor was "Distractability."

Another factor had thirteen items loading on it (in the range of 0.61 to 0.36). The gist of these is harder to see at first, by comparison to the Distractability factor, but think about these: "I reread my notes on the passage"; "After having already read the passage carefully one time, I reread the complete passage carefully at least one more time"; "I test myself by going over the passage page by page and trying to memorize the important points and details." This is surely a "processing" variable, or a "primary strategy," but notice that if carried to extreme, it would be counter-productive. A factor here should be thought of as something people do more or less of, and in this case more is probably bad while a moderate amount is probably good, at least in the studying for homework condition. "Industry/Compulsivity" was chosen as a label for this factor with the idea that industry carried to extremes is compulsion.

A third factor had but five items loading on it (in the range of 0.48 to 0.35). They were these: "I try to understand the text well enough to put it in my own words without losing the author's meaning"; "I test myself by trying to put the main points of the passage in my own words"; "I try to relate the writer's ideas to things I learned in other places"; "I try to judge how the writer's ideas make sense"; "I underline when I read." The striking thing about these items is their opposing nature in comparison to compulsion, which was marked by memorization. Perhaps this reflects a bias toward Eriksonian psychology but the name picked for this factor was "Initiative," in the sense of to take initiative with. Perhaps the factor name should be "Personalizing," but that seemed less precise. In relation to other study skills taxonomies, in any case, this is definitely another "process" variable or "primary" strategy.

In summary, here is the net result of our factor analysis of data gathered on the studying for homework condition: the strongest study habits factor was "Distractability," similar in nature to the factor others have called state variables or support strategies. Distinct from that, however, a factor we called "Industry/Compulsivity" emerged for this studying condition. And, in addition to that, there emerged a factor which was labeled "Initiative."

Consider now the two factors which emerged for the studying for a test condition:

One factor seems to carry across both conditions and that is "Distractability." The average loadings of the seven key items is the same in each condition (0.58). This factor seems to operate whether people are studying for homework or studying for a test.

The other two factors do not reappear; they are supplanted by a single new factor for this different condition of study. Several items previously loading on the Initiative factor ("I underline," "I try to put the main points in my own words") now appear with items previously comprising the Industry factor. The change in definition of condition is that now the condition of study is much more specific-goal oriented. It seems that in studying for homework, students are more likely to take clear initiative, perhaps in order to establish a goal (e.g., "to relate the writer's ideas to things I learned in other places"). But when they are studying for a test, the goal is prescribed, and it's not initiative but compliance which is demanded. This factor, therefore, is not Industry/Compulsivity or Initiative but "Assimilation," in the ordinary sense of that term, "to absorb and incorporate." It is similar to the factor others have called "encoding." In fact, the data verify Anderson and Armbruster's distinction between "encoding" and "focus of attention."

The third question remains, the one concerning the possible difference among study habits of students at different grade placements. The shortest answer to the question is, "Yes, they differ but no, not as much as you might think." For the Distractability factor, there were no differences—students report approximately the same abilities to concentrate no matter what their grade level, seventh, tenth, or college, and this holds for both homework and studying for a test conditions. For the Initiative factor, applicable only to the studying for homework condition, there were only slight differences among the age groups, but the differences were in the direction one might expect: college readers tend to have more study initiative than do tenth graders who in turn have somewhat more initiative than seventh graders. For the factor called Assimilation, applicable only to the studying for a test condition, there was a difference in favor of tenth graders and college students, taken as a group, by comparison to seventh graders. This factor may level off in development somewhere between seventh and tenth grades. For the Industry/Compulsivity factor, however, the differences among groups are not as straightforwardly interpretable, though the differences were more dramatic than those obtained for any other factor. The highest Industry/Compulsivity score was for seventh graders, with a significantly lower score for tenth graders, and a significantly even lower score for college age readers. On close inspection, it is apparent that the younger readers were scoring highest on precisely the several items which would define industry as compulsivity, items related to careful rereading and attempts to memorize. What this suggests is that younger readers' study habits are not as wrong as they are inefficient, and it's not as if younger readers don't know what to do as much as they do during a homework assignment what older students save for review sessions. Most of the items loading on this factor refer to practices which would be effective if applied judiciously and selectively but which would be counterproductive if applied indiscriminately.

The obvious next step in this research, which it is hoped others will join us in undertaking, is to administer this revised scale to groups of students at times immediately following homework assignments and preparation for a test. In each case, students will be asked whether they have done the things the items refer to. Then the question of whether the results of the survey bear any relationship to performance on homework tasks or test performance will be asked.

Attached to the survey, we had several "extra" questions about study habits, one of which was "How did you develop your present methods of study?" The answer was consistent enough to give pause—"I more or less developed my present study methods on my own, with little or no help from teachers or books on how to study." Studying is too important a skill to be left to chance. Perhaps with a better understanding about what constitutes study and how it is related to school performance, some of the chance can be eliminated.

REFERENCES

Anderson, T. H., & Armbruster, B. B. *Studying* (Center for the Study of Reading, Tech. Rep. 155). Champaign, Ill.: University of Illinois, January, 1980.

Dansereau, D. F. The development of a learning strategies curriculum. In H. F. O'Neil, Jr. (Ed.), *Learning strategies*. New York: Academic Press, 1978.

Wrenn, C. G. *Study habits inventory*. Stanford, California: Stanford University Press, 1934/1941.

JANE HANSEN
University of New Hampshire

GEORGANNA AHLFORS
University of Wisconsin-River Falls

INSTRUCTION IN INFERENTIAL COMPREHENSION: AN EXTENSION AND A SUMMARY

Readers actively construct meaning for a text when they infer relationships between the text and their stored knowledge (Collins, Brown & Larkin, 1980). Even though comprehension depends on inferential thinking, teachers seldom teach their students how to draw inferences (Durkin, 1978–79; Neilsen, 1981; Palmer, 1981). Instructional research could narrow the gap between comprehension research and the real-world variables of students, materials, and teachers.

This article presents the third study in a series of instructional studies designed to improve comprehension instruction by teaching students how to draw inferences. In the first study, second-grade, good readers received ten weeks of instruction intended to influence their comprehension of stories in their basal readers (Hansen, 1981b). Two types of instruction were tried: (1) pre-reading, strategy discussions, and (2) guided-reading, inferential discussions. Both instructional procedures significantly improved the children's inferential comprehension. In the second study fourth-grade, good and poor readers received similar instruction. In this case, the instruction benefited the poor readers but had no effect on the comprehension of the good readers (Hansen & Pearson, 1982). The present study with fifth-grade, good readers extended the two previous studies.

It provided an additional test of the method with upper-elementary, good readers. We used more difficult instructional materials than we had used with the fourth-grade, good readers because they used fourth-grade basals and found these stories so easy they didn't need to use the new strategy. This time we continued to use basal-reader selections but we used only expository selections and we chose selections from sixth-grade basals. Thus, the students read articles written at their instructional reading level rather than articles written for their grade level.

Also, we trained the teachers to write the pre-reading discussion questions. We found in the fourth-grade study that teachers could conduct the discussions but they followed scripts provided by the researchers. However, we want to suggest this method to teachers so we needed to find out whether teachers could write the lessons.

Finally, we extended the pre-reading, strategy discussions into the area of vocabulary development because the vocabulary instruction teachers provide doesn't improve comprehension (Ahlfors, 1979). For the original strategy discussions, we selected three important ideas from the passage to be read and wrote two discussion questions for each of the ideas (Hansen, 1981a). The first question required the students to discuss prior knowledge relevant to the topic and the second question required students to hypothesize what form similar information might take in the passage. This example was used with *The Moon Colony*, Ginn, 1976:

Important Idea:
Early moon colonists will find life difficult.
Prior knowledge question:
Tell about a time you had to "rough it."
Hypothesis question:
In the next story how do you think the moon colonists will have to "rough it"?

The teacher who wrote the vocabulary introductions followed the above format but he chose three important vocabulary words instead of three important ideas. Several criteria guided the selection of the words: (1) they represented important concepts in the article, (2) the teacher thought the words were unfamiliar to the students, (3) they fit the prescribed format, and (4) they were potential additions to the students' personal vocabularies. He used this example as part of the introduction to *Meat-Eating Plants*, Rand McNally, 1978:

Vocabulary word: thrive
Prior knowledge question:
To thrive means to flourish or live happily and grow well. Monkeys thrive on bananas. What kinds of food do you thrive on?
Hypothesis question:
In the article the plants thrive on something unusual. What do you think it might be and why would it be unusual?

METHOD

Subjects

We conducted the study in one school in a middle-class community of Dover, New Hampshire, a town of approximately 20,000. We used Stanford Achievement Test scores and teacher judgment to determine which of the 100 fifth-grade students in the school qualified as good readers. From 46 students who qualified, we randomly selected 33 and randomly assigned each of them to one of three groups, with 11 per group. The 33 students all had SAT scores within the range of 5.2–10.2 (grade-equivalent scores). The means for the three groups were statistically equivalent: 7.22, 7.24 and 7.25. We randomly chose the three teachers from five available teachers and randomly assigned each teacher to one of the three groups.

Instructional Materials

We used photocopied sets of expository passages from various current basal series. The teachers also received the accompanying teacher's guide to use at their discretion on the days when activities connected with the study did not occur.

Procedure

The study lasted for ten weeks and the students read one passage each week.

We established two experimental groups and a control. The first experimental group, the Important-Ideas Group, received the instruction used in the fourth-grade study which included both the pre-reading discussions described above and post-reading discussions. The post-reading discussions consisted of inferential questions that required the students to go beyond the information in the passage to answer the questions (Hansen, 1981b; Hansen & Pearson, 1982). The second experimental method, the Vocabulary Group, received the vocabulary version of the important-ideas discussions described above (and the post-reading discussions). These vocabulary discussions were the only pre-reading instruction provided for this group, whereas the Important-Ideas Group received traditional vocabulary introductions in addition to the important-ideas introductions.

Also, both experimental groups took part in a strategy-awareness discussion prior to the above story introductions. In the strategy-awareness discussions the students talked about situations when they read independently and used the inferential-thinking strategy; they told about times when they related something they were reading to their own lives.

The control group received traditional vocabulary instruction, content introductions as prescribed in the basal manuals, and post-reading discussions of 80% factual questions and 20% inferential questions.

The teachers taught one practice lesson with their group before the actual study began. Throughout the study one of two researchers either observed the teachers weekly or video taped or audio taped them. A conference was held with each teacher after an observed or taped session. We monitored the teachers this way rather than switching them midway in order to make this study as similar to a real classroom experience as possible because, in the fourth-grade study, the teachers had said it isn't natural for teachers to switch methods and groups every five weeks.

Dependent Measures

We used two dependent measures. First, each student completed a nine-question worksheet of open-ended questions after the discussion of each of the ten instructional stories. There were three factual questions, three inferential questions, and three vocabulary questions. The factual questions were rewordings of sentences lifted from the text and the inferential questions followed the format of the inferential discussion questions (Hansen, 1981b; Hansen & Pearson, 1982). The vocabulary questions were inferential. For example, we asked this vocabulary question for the word naturalist: Pretend some naturalists lived in Dover. Name a specific place where they might often study.

The second dependent measure was a set of three posttests administered to groups at the conclusion of the study. The students read a set of three expository, basal-reader passages, one at a sixth-grade reading level (the reading level of the instructional materials), one at a seventh-grade reading level and one at a fifth-grade reading level. After each selection (one on each of three days, administered in random order), the students completed a twelve-item test of open-ended questions, with six inferential items and six factual items.

RESULTS

First, the students in the Important-Ideas Group did not receive significantly higher scores on either dependent measure than the students in the Control Group. (See Table 1 for all results.) Second, the teachers could write the discussion questions for both of the introductions. Third, the students in the Vocabulary Group performed as well as the students in the Important-Ideas Group.

TABLE 1
Worksheet and Posttest Data
MANOVA, Means and (Standard Deviations)

*Worksheets

**MANOVA *F* (9,93) = 1.74, $p > .05$

	Literal Probes	Inferential Probes	Vocabulary Probes
Vocabulary Group	13.82 (3.74)	17.99 (2.77)	18.00 (2.37)
Important-Ideas Group	14.64 (3.17)	17.36 (3.78)	15.64 (3.78)
Control Group	12.82 (4.45)	14.55 (4.28)	14.46 (5.07)

Posttests

Fifth-grade passage
MANOVA *F* (6,78) = 1.02, $p > .05$

	Literal Probes	Inferential Probes
Vocabulary Group	3.09 (1.58)	3.00 (1.27)
Important-Ideas Group	3.36 (1.50)	2.73 (1.49)
Control Group	2.27 (.91)	1.82 (1.33)

Sixth-grade passage
MANOVA *F* (6,78) = .33, $p > .05$

	Literal Probes	Inferential Probes
Vocabulary Group	1.91 (1.22)	2.55 (1.21)
Important-Ideas Group	1.73 (1.68)	2.27 (1.19)
Control Group	2.00 (1.27)	2.36 (1.43)

Seventh-grade passage
MANOVA *F* (6,78) = .71, $p > .05$

	Literal Probes	Inferential Probes
Vocabulary Group	2.00 (1.00)	2.36 (1.03)
Important-Ideas Group	2.00 (1.55)	1.55 (.52)
Control Group	1.64 (1.36)	2.00 (1.73)

*Worksheet scores were summed across all ten weeks.

**The MANOVA statistics reflect four instructional groups. The fourth group was not mentioned in this article due to lack of space. If this group had not been included in the original analyses, the significance of the results would not have changed.

CONCLUSIONS

This line of research explored the feasibility of improving elementary students' ability to draw inferences. The studies were based on the assumption that in order to understand new information, comprehenders must relate the new information to some other information they already understand. Mature readers do this. And children do this in nonreading situations. However, children do not spontaneously draw these inferences when they read. Therefore, this instructional method was originally developed to teach children to apply an inferential thinking skill they already possessed to reading situations.

However, this method did not influence the comprehension of fifth-grade, good readers. The students found the passages so difficult that these introductions and discussions did not provide enough help. The expository passages were typical expository passages with heavy concept loads. We should have modified the situation when it became apparent the students didn't comprehend the stories well. The means reveal low scores for good students; a perfect score for each set of worksheet probes would have been 30 and these students got less than two-thirds correct. In retrospect, we did not have success with either fourth- or fifth-grade good readers because the instructional materials were inappropriate. Therefore, it remains to be seen whether these discussions will work for upper-elementary, good readers.

On the other hand, because we have tested this teaching procedure in several contexts, we can suggest it to teachers of certain types of students. Good readers in primary grades and poor readers in intermediate grades found these introductions and discussions helpful; and, further, the poor readers used these thinking skills in independent reading situations so the method is especially helpful for them.

The teachers could write the questions for both of the pre-reading discussions. The teacher who wrote the introductions based on the vocabulary words found his questions easier to write than did the teacher who wrote the introductions based on three important ideas. After the teachers had written three sets of questions, they demonstrated competence with the task. It took them twenty minutes to write each introduction. Thus, this study provides us with support when we advocate this method to teachers.

The vocabulary variation of the pre-reading discussions adds flexibility to the original method of discussing important ideas. The words represent important ideas, or concepts. Teachers can now select important words and/or important ideas as the basis for introductions.

Finally, a strong motivation factor has emerged while working with teachers on these projects. This is important because, unfortunately, many teachers do not like to teach reading. One of the teachers who participated in this study belonged to that category. When we first proposed the study to him, he responded, "Teaching reading is the most boring part of my day and reading class is dull for my students." However, as the study progressed, the tone of his comments changed. The students enthusiastically took part in these discussions and this, in turn, stimulated him. A visit to his classroom several months after the study revealed that his enthusiasm was still high, "This is a rewarding way to teach!"

ACKNOWLEDGEMENTS

This study was funded by CURF grant No. 827 from the University of New Hampshire. We want to thank Ruth Hubbard, the research assistant. Also, we want to thank the personnel of the participating school: Paul Curtis, John D'Vittori, Dot Colprit, Cindy Harmon, and Art Bettencourt.

REFERENCES

Ahlfors, G. *Learning word meanings: A comparison of three instructional procedures.* Unpublished doctoral dissertation, University of Minnesota.

Collins, A., Brown, J. S., & Larkin, K. M. Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum, 1980.

Durkin, D. What classroom observations reveal about reading comprehension. *Reading Research Quarterly*, 1978–79, *14*, 481–533.

Hansen, J. An inferential comprehension strategy for use with primary grade children. *The Reading Teacher*, March 1981, *34*, 665–669 (a).

____ The effect of inference training and practice on young children's comprehension. *Reading Research Quarterly*, 1981, *16*, 391–417 (b).

Hansen, J., & Pearson, P. D. *An instructional study: Improving the inferential comprehension of fourth-grade, good and poor readers* (Teach. Rep. No. 235). Champaign, IL: University of Illinois, Center for the Study of Reading, March 1982.

Neilsen, A. R. *Conditions affecting reading to learn in content area classrooms*. Paper presented at National Reading Conference, Dallas, December 1981.

Palmer, W. *Teacher behavior in relation to reading comprehension*. Presentation at University of New Hampshire, December 1981.

Rand McNally Reading Program. *Shifting Anchors* (Level 13). Chicago: Rand McNally & Company, 1978.

Reading 360. *Measure Me, Sky* (Level 12). Lexington, Massachusetts: Ginn and Company, 1976.

DAVID A. HAYES
JOHN E. READENCE
University of Georgia

EFFECTS OF CUED ATTENTION TO ILLUSTRATIONS IN TEXT

There is a commonsense belief that illustrations in texts serve as aids to comprehension. Certainly where their intended purpose is to be informative, rather than merely decorative, illustrations are supposed to amplify the print with additional visual information that can clarify the content for readers. Illustrations are thought to be especially helpful if the content cannot be efficiently dealt with through a strictly verbal medium as, for example, content involving three-dimensional spatial relations (Schallert, 1980). Whether illustrations in texts actually serve this end and help readers understand their texts cannot be determined from the existing research on visual aids to reading.

In spite of a sizable body of research in this area, conclusions as to the efficacy of adjunct illustrations remain tentative with respect to questions of text design and instructional practice (Readence & Moore, 1981; Rigney & Lutz, 1976; Schallert, 1980). For the most part, investigations on visual aids to reading comprehension suffer from generality of purpose and artificiality in procedure. Typically, they pose very broad questions about the effectiveness of texts' illustrations; yet, just as typically, they fail to inquire into possible circumstances under which illustrations might prove helpful. Only a handful of studies have been undertaken with school-age children who actually read the experimental materials for themselves, rather than listened to them and that used materials representative of traditional school texts rather than programmed materials, filmstrips, or cartoons (Haring & Fry, 1979; Holliday, 1975; Koenke & Otto, 1969; Rasco, Tennyson & Boutwell, 1975; Thomas, 1978). Of these studies none, except for the one conducted by Rasco, Tennyson & Boutwell (1975), directed subjects to attend to the illustrations in the text. None investigated students' use of illustrations in study tasks like those encountered in everyday classroom activities, e.g., problem-solving tasks which entail referring to the text for critical information. And none controlled for the effect of specific text information, which may preclude generalizing research findings from subjects' encounter with a single text to their performance on any text (Clark, 1973).

Given the limitations of the previous research on visual aids to reading, the widely accepted conclusion among researchers (such as Concannon, 1975; Samuels, 1970; Vernon, 1954) that illustrations used as adjuncts to printed text do not facilitate comprehension hardly seems warranted, especially when such a conclusion runs contrary to our intuitions about the clarifying power of illustrations, even for children. Thus, the present investigation was undertaken with the limitations of the previous research in mind.

The purpose of the study was to address these limitations as we sought to find out not only whether illustrations would help children understand prose material but whether giving them oral instructions to look at the illustrations would make a difference. Rather than following the usual procedure of giving students a text either with or without pictures and then giving them a test on it, we attempted to determine the effect of illustrations in text under circumstances more likely to be found in classrooms, i.e., in a study situation in which students have text materials available to them to deal with study problems presented either in the text or in an accompanying workbook. In other words, the dependent measure was not posttest or recall performance; it was, rather, students' success at working study problems with the text available.

The specific questions of the study were:

1. Would augmenting text with illustrations make a difference on students' success with associated problem-solving tasks in a classroom setting?
2. Would directing students to view the illustrations in text make any difference in these tasks?
3. In the absence of illustrations in text, would instructing students to imagine their own mental pictures make a difference?
4. Would any of these conditions make a difference for any of four levels of reading ability ranging from below average to above average reading ability?

METHOD

The subjects were 82 eighth-graders in a middle school in rural northeastern Georgia. The subjects read and did study exercises on four different texts under four different treatment conditions.

Four different texts were used in order to control for the effect of specific text information. The texts were 300-word selections on the subject of simple machines. The topics of the four texts were: Mechanical Advantage, Levers, Pulleys, and the Inclined Plane. A 10-item study exercise was included with each selection. The selections and their accompanying study problems were taken from a widely used science textbook by Heimler & Neal (1975). These texts were assembled into packets so as to provide four different sequences of topic presentation. The four text sequences were ordered so that all texts were comparably distributed across all treatment conditions. For the experiment, subjects were assigned at random to one of the four text sequences.

To introduce subjects to the experimental activity, the teacher read aloud the cover page of their packets while the subjects read it silently. It was a 200-word general statement emphasizing the usefulness of simple machines, especially for the development of complex machines, and identifying six types of simple machines.

Under the first treatment condition the subjects read a text without illustrations and performed the study exercises without any instructions other than to refer to the text in order to do the problems.

After all subjects completed this task, they were given the task of the second treatment condition, which was like the first condition except this time the texts were augmented by illustrations. The illustrations were line drawings which depicted applications of the text content. No instructions were given to look at these drawings, only to refer to the text in responding to the accompanying problems.

For the third treatment, subjects turned to their packets' third text which was also augmented by line drawings. Again the subjects were told to read the text and do the accompanying problems, but this time they were told that the pictures should help them understand the material and advised that they should therefore pay careful attention to the pictures.

For the fourth treatment, the subjects turned to the last text in their packets.

This text had no illustrations, as in the first treatment condition. This time, however, subjects were instructed to try to imagine their own pictures to go along with the materials.

RESULTS AND DISCUSSION

The treatment conditions across four levels of reading ability were compared by a repeated measure analysis of variance. No statistically significant interactions between any treatment condition and reading level were observed. A significant main effect for reading level was observed, $F(3,3)=13.61$, $p<.001$. As expected, Scheffé post hoc tests showed that above average readers performed better ($\bar{X}=6.37$) than average and slightly below average readers ($\bar{X}=4.89$ and $\bar{X}=4.64$, respectively), who in turn performed better than below average readers ($\bar{X}=3.54$). A significant main effect was also observed for treatment condition, $F(3,78)=11.26$, $p<.001$. Scheffé post hoc tests showed that subjects performed better on texts augmented with illustrations ($\bar{X}=5.21$ and $\bar{X}=5.62$), than on texts without illustrations ($\bar{X}=3.72$ and $\bar{X}=4.12$). No effect was noted for instruction condition, either to attend to illustrations when they were present or to imagine the content when they were not present.

The data indicate that illustrations do facilitate comprehension of printed text for upper elementary level children, at least with materials which require comprehension of spatial relations. That no significant difference was observed between the first no-illustration condition and the final no-illustration condition diminishes the possibility that the result was due to an order effect. This finding is significant in light of previous researchers' failure to detect the effect of illustrations on children's reading comprehension. The present finding confirms classroom teachers' everyday observations and commonsense conviction that informative illustrations in text help children comprehend and learn from printed language. The finding is also significant in terms of the study's ecological validity. Subjects responded to experimental materials adapted from a textbook widely used in science classes, and the experimental activity closely resembled the kind of study activity which commonly takes place in upper elementary and secondary classrooms.

Although this study found no significant effect stemming from instructions to attend to the illustrations in the text or, in their absence, to visualize the content, sound instructional practice nevertheless supports such direction from the teacher. That this study failed to detect an effect for these instructions hardly means that there is no effect for this kind of direction. Hopefully, future research which improves upon the present effort will prove useful for both classroom teachers and publishers who share the mutual goal of helping students learn from text.

REFERENCES

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335–359.

Concannon, S. J. Illustrations in books for children: Revising research. *The Reading Teacher*, 1975, *29*, 254–256.

Haring, M. J., & Fry, M. A. Effect of pictures on children's comprehension of written text. *Educational Communication and Technology Journal*, 1979, *27*, 185–190.

Heimler, C. H., & Neal, C. D. *Principles of science, book one* (3rd ed.). Columbus, Ohio: Merrill, 1975.

Holliday, W. G. The effects of verbal and adjunct pictorial-verbal information in science instruction. *Journal of Research in Science Teaching*, 1975, *2*, 77–83.

Koenke, K., & Otto, W. Contribution of pictures to children's comprehension of the main idea in reading. *Psychology in the Schools*, 1969, *6*, 298–302.

Rasco, R. W., Tennyson, R. D., & Boutwell, R. C. Imagery instructions and drawings in learning prose. *Journal of Educational Psychology*, 1975, *67*, 188–192.

Readence, J. E., & Moore, D. W. A meta-analytic review of the effect of adjunct pictures on reading comprehension. *Psychology in the Schools*, 1981, *18*, 218–224.

Rigney, J. W., & Lutz, K. A. Effect of graphic analogies of concepts in chemistry on learning and attitude. *Journal of Educational Psychology*, 1976, *68*, 305–311.

Samuels, S. J. Effects of pictures on learning to read, comprehension, and attitudes. *Review of Educational Research*, 1970, *40*, 397–407.

Schallert, D. L. The role of illustrations in reading comprehension. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.). *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education*. Hillsdale, N.J.: Erlbaum, 1980.

Thomas, J. L. The influence of pictorial illustrations with written text and previous achievement on the reading comprehension of fourth grade science students. *Journal of Research in Science Teaching*, 1978, *15*, 401–405.

Vernon, M. D. The instruction of children by pictorial illustration. *British Journal of Educational Psychology*, 1954, *24*, 171–179.

LEA M. McGEE
Louisiana State University

THE INFLUENCE OF METACOGNITIVE KNOWLEDGE OF EXPOSITORY TEXT STRUCTURE ON DISCOURSE RECALL

A major emphasis in reading programs is helping students acquire strategies for learning and remembering information from reading. In recent years, research has shown that following the author's structure to guide encoding and retrieval of text material is an effective reading strategy. This strategy, not surprisingly referred to as the "structure strategy" (Meyer, Brandt, & Bluth, 1980), not only increases the likelihood that important ideas will be recalled, but also enhances the amount of information recalled. Readers using the structure strategy search for the author's thesis. In addition, they search for specific kinds of relationships which tie thesis statements together as well as link these statements to other ideas in the passage. Retrieval proceeds as a top-down search for text information and is guided by the hierarchy formed by thesis ideas and relationships among those ideas and other ideas in the passage.

Adults are fairly adept using the structure strategy. Several studies (Meyer, 1975; Meyer & Freedle, Note 1; Meyer, Rice, Knight, & Jessen, Note 2; Taylor, 1980) have shown that more than 80% of adults at all ages use the author's structure to organize text recall. However, fewer young students appear to use this strategy. Meyer, Brandt, and Bluth (1980) showed that the percentage of older students using the author's structure to organize recall who were good readers approached the percentage of adults (72% of 32 ninth graders). However, much lower percentages of average and particularly poor readers (57% and 12% each of 35 ninth graders) used the author's structure to organize their recall. Similarly, Taylor (1980) and McGee and Niles (1981) found that only 57% of sixth grade good readers (N=17) and 60% of fifth grade good readers (N=20) used the author's superordinate-subordinate structure to organize recall. In order to provide information for helping students apply more effective reading strategies, the purpose of the present study was to explore why young students seem to be less likely to use the structure strategy than adults.

There may be at least two reasons why adults appear to be more likely to use the structure strategy than young students. In previous studies (Meyer, Brandt, & Bluth, 1980; Meyer & Freedle, Note 1; Taylor, 1980) some students and adults read and recalled the same passages. A higher percentage of adults may have used the author's structure because they were reading relatively easier passages. Therefore, they may have been able to allocate more cognitive capacity to noting and following the author's structure during reading and recall. In these studies, young readers' ability to follow text structure may have been masked because they were reading relatively more difficult text written at or above their reading ability level. The present study investigated whether young students are as adept as adults in using text structure to organize their recall when they too are given relatively easy text.

The second reason adults are more likely to use the structure strategy than young students may be related to their metacognitive knowledge of text structure. Metacognitive knowledge of structure, in contrast to spontaneous use of knowledge of structure to organize recall, refers to a more conscious awareness of how structure is used in text-related tasks (Vygotsky, 1934/1962). For example, readers with metacognitive knowledge of structure would be likely to not only follow structure when it was present, such as when reading and recalling, but also to produce structure when it was not available to follow, such as when writing a novel composition. Research has demonstrated that young students are at least sensitive to structural aspects of text, using structure to organize their recall (Taylor, 1980; Meyer, Brandt, & Bluth, 1980; McGee & Niles, 1981). Further, this sensitivity influences recall. Metacognitive knowledge of structure has also been shown to correlate with differences in recall (McGee, 1980; McGee & Niles, 1981). Causal relationships between recall and metacognitive knowledge of structure have also been found (Bartlett, 1978; McDonald, 1978; Armbruster, 1979). None of these studies have examined what the respective influence of sensitivity to structure (spontaneous reproduction of structure in recall) and metacognitive knowledge of structure (more conscious creation of structure in novel compositions) is on predicting recall differences. It would be expected that readers who have a metacognitive knowledge of structure would be more likely to use the structure strategy more effectively and recall text better than readers who are merely sensitive to text. Moreover, it would be expected that more adults would be likely to have metacognitive knowledge of structure than young readers. Therefore, the purpose of the present study was to examine adults and students' metacognitive ability to create structure in novel compositions. In addition, the relationship between spontaneous sensitivity to structure as well as the influence of metacognitive knowledge of structure on amount of information in recall was investigated.

METHOD

A total of fifty-five students were randomly selected from three groups of average readers at the sixth grade, ninth grade, and college undergraduate levels. Twenty-one sixth graders whose reading grade equivalent scores were between 5.3 and 6.7 on the SRA Standardized Achievement Test were chosen from the sixth grade population of an elementary school. Sixteen ninth graders scoring between 8.1 and 9.8 on the reading subtest of the Stanford Achievement Test were selected from the ninth grade population at a junior high school. Eighteen undergraduate readers scoring between the 38th and 63rd percentile on the Nelson Denny Reading Test were chosen from four undergraduate education classes at a university.

Three texts were written at the fourth, seventh, and tenth grade readability levels (Fry, 1968). These texts had approximately 200 words each and contained a problem/solution top-level structure (Meyer, Note 3). Each passage was segmented into idea units which were placed in a hierarchy using an analysis of discourse procedure (Taylor, 1978). The topics of the passages read by sixth grade, ninth grade, and adults, respectively, were forest fires, oil shortages, and inflation. Figure 1 displays the passage recalled by sixth graders.

The three passages were equated for familiarity using a random sample of ten subjects at each age level who were not participating in the study. These subjects were orally asked five questions about the passage information without having read the passage. There were no significant differences among the mean number of questions correctly answered by each of these groups (1.78, 2.04, and 1.95).

In order to stimulate production of compositions, a topic for discussion at

FOREST FIRES

An important problem for forest rangers is how to protect forests from being ruined by fire. Each year thousands of trees are destroyed by fire. Careless campers do not put their fires completely out. Cigarettes are left to burn on the dry ground. These small fires in dry forests can burn thousands of trees. One solution to the forest fire problem is to man lookout stations and use helicopters to spot fires. Fires that are spotted right away can be put out before they get too big to handle. Then fires will cause less damage to the forest. A second solution to the problem is to have experts and bulldozers ready to move in quickly to fight the fire. Bulldozers can throw huge amounts of dirt on a fire in a short time. The dirt helps put the fire out quicker. A third solution is to build fire lanes in the forests. Fire lanes are long breaks in the forest where there are no trees. These breaks prevent the fire from spreading and getting too large.

FIGURE 1: Passage Recalled by Sixth Grade Students.

each age level was selected after consulting with the students' teachers. The topics selected were those that teachers felt students would know most about from recent study: education in India, storing nuclear wastes in Louisiana, and handling discipline problems.

One half of the subjects were randomly chosen to complete the composition task first followed by the recall two weeks later. The procedure was reversed for the other group. In the composition condition, groups of 9–11 subjects participated in a 20–30 minute discussion led by the investigator. The discussion consisted of subjects giving descriptions and explanations of the problems and solutions associated with the topic. These points were summarized by subjects as well as by the investigator at least three times during the discussion. The words "problem" and "solution" were emphasized. Then subjects were instructed to write a summary of the discussion. They were reminded that the compositions should reflect their best effort.

In the recall condition, subjects were tested in groups of 5–6. They received a booklet containing a practice passage and the recall passage, each passage followed by a blank page. They were instructed to read the practice passage and turn to the next page. On that page they were told to write everything they could remember about the passage without looking back. The procedure was repeated for the recall passage.

Recall protocols were scored for the percentage of idea units recalled using Taylor's (1978) scoring procedure. The reliability coefficient between two independent raters was .91. A weighted central idea score was calculated for use in the regression analysis. Thesis statement ideas located at the top level of the text hierarchies were credited with 1 point. Major ideas located at the second level were credited with two points. Detail ideas located below the second level were credited with half a point. The weighted recall score was employed in the regression analysis because the procedure used in analyzing the structure found in recall created a bias between type of structure in recall and level of idea unit in recall. Students had to recall most of the ideas located on the top level of the text hierarchy in order to have their recall classified as having a problem/solution structure. The weighted score was used to reduce this bias while still maintaining an emphasis on recall of important ideas.

Recall and composition protocols were analyzed to determine how well information was organized into a problem/solution structure. Meyer, Brandt, and

Bluth's (1980, pp 84–85) ten category system was used with one exception. Recalls and compositions that had clusters of ideas only relating to the problem were credited equally as recalls and compositions with clusters of ideas only related to the solution. Then, recalls and compositions were further classified into three categories. Recalls and compositions that had two clusters of related ideas (one related to the problem and one related to the solution) as well as having the words "problem" and "solution" explicitly stated were identified as having a fully developed problem/solution structure. Recalls and compositions that had two related clusters of ideas but had only the word "problem" or "solution" explicitly stated or neither were identified as having a less well-developed problem/solution structure. Finally, recalls and compositions that had only one cluster of ideas, presented ideas in a list-like fashion, or that had another structure were identified as not having a problem/solution structure.

RESULTS

The percentage of recall (not weighted) for sixth grade, ninth grade, and adults was .322, .590, and .591, respectively. An analysis of variance revealed a significant difference among the mean recall scores of the three groups, $F(2,52)=56.58$, $p<.01$. Tukey's post hoc analysis demonstrated that sixth graders performed less well than ninth graders or adults ($p<.01$).

Tables 1 and 2 present the percentage of subjects who used a fully developed, less well-developed, and no problem/solution structure in recall and composition. Chi-square analyses were conducted to determine whether there were differences among the three groups in the kind of structure used to organize recall and compositions. Table 1 demonstrates that in recall more sixth graders tended to use the less well-developed problem/solution structure while more ninth graders and adults tended to use the fully developed problem/solution structure, $\chi^2_4=89.517$, $p<.001$. Table 2 shows that in compositions more sixth graders tended not to use a problem/solution structure, more ninth graders tended to use the less well-developed problem/solution structure, and more adults tended to use the fully developed problem/solution structure, $\chi^2_4=73.418$, $p<.0001$.

TABLE 1
Percentage of Students and Adults Who Use Structure in Recall

Groups	Fully Developed Problem/Solution Structure	Less Well-Developed Problem/Solution Structure	Structure Other Than Problem/ Solution
Sixth Grade Students (N = 21)	19.1%	61.9%	19.0%
Ninth Grade Students (N = 16)	75.0%	18.8%	6.2%
Adults (N = 18)	83.3%	11.1%	5.6%

$\chi^2_4=89.517$, $p<.001$.

TABLE 2
Percentage of Students and Adults Who Use Structure in Composition

Groups	*Fully Developed* Problem/Solution Structure	*Less Well-Developed* Problem/Solution Structure	Structure Other Than Problem/ Solution
Sixth Grade Students (N = 21)	4.8%	23.8%	71.4%
Ninth Grade Students (N = 16)	25.0%	43.8%	31.2%
Adults (N = 18)	50.0%	33.3%	16.7%

$\chi^2_4 = 73.418$, $p < .0001$.

A regression analysis was used to determine whether use of structure in recall and composition influences recall of important information. Use of a problem/ solution structure in recall explained 40.8% (F = 108.25) of the variance in weighted central idea recall. Use of a problem/solution structure in composition explained an additional 10.5% (F = 46.49) of the variance.

DISCUSSION

The purpose of this study was to investigate adults' and students' ability to use a problem/solution structure to organize recall when reading relatively easy text. Also investigated was the use of structure to organize composition and the influence of these two abilities on recall of major text ideas. The results suggest that giving students relatively easy text significantly influences how well they are able to use text structure in organizing recall. The percentage of sixth graders (81%) who used either the fully or less well-developed problem/solution structure in recall was similar to that of the ninth graders and adults (94%). When given relatively easy text, the performance of even sixth grade average readers in the present study was better than the performance of both good sixth and ninth grade readers in previous studies (Meyer, Brandt, & Bluth, 1980; Taylor, 1980) and as good as adults in other studies (Meyer & Freedle, Note 1; Taylor, 1980). Sixth graders may have used less cognitive capacity for decoding text (Britton, Glynn, Meyer, & Penland, 1982) and, therefore, have had more capacity for noting and remembering text structure.

Developmental trends are apparent in students' ability to use structure to organize compositions and in the amount of central information recalled. Sixth graders recalled less information than either ninth graders or adults. Moreover, only 29% of the sixth graders used either the fully or less well-developed problem/solution to organize their compositions. In contrast, 69% and 94% of the ninth graders and adults employed these structures in their compositions. Sixth graders may have recalled less information than older readers because they are less adept at applying their knowledge about structure. Although 81% of the sixth graders used a problem/solution structure to organize their recall, only 29% of these readers used this structure in their composition. These students may have intuitively followed structure during reading, but did not use knowledge about structure to produce compositions. This result supports the notion that younger students may be less effective in using the structure strategy because they have less metacognitive knowledge about structure.

Results of the regression analysis confirm the conclusions of earlier studies that the ability to use structure in text is strongly related to how well readers recall information. However, in the present study the variance accounted for by using structure to organize recall may be spuriously high. The weighted recall score included points for top level thesis statements which had to be included in order for recalls to be classified as having a problem/solution structure. Therefore, the variance accounted for by structure in recall may have been inflated. However, there was no direct relationship between the recall measure and structure used in composition. The 10.5% additional variance accounted for by structure in composition demonstrates how powerful knowing about and applying knowledge about structure is in the comprehension process.

In conclusion, students may best be able to apply effective reading strategies when processing easier text. Although no conclusions about causation can be drawn from the results of the present study, it seems that teaching students to write well-organized text may be important in developing effective reading strategies. However, caution must be taken in drawing this conclusion. This study examined the ability to produce ony one "ideal" text structure—the problem/solution rhetorical structure as defined by Meyer (Note 3). This structure is ideal in the sense it is assumed to produce optimal recall. However, Britton, Burgess, Martin, McLeod, and Rosen (1975) maintain that there are many "ideal" rhetorical structures depending on the purpose of writing or reading. Teaching students to create structures based on only one rhetorical grammar would be an over-simplification of the implications of the present study. Rather, helping students write in any more organized fashion to suit their various purposes in composing may help them learn to note and use many different organizational patterns as they are reading.

REFERENCE NOTES

1. Meyer, B., & Freedle, R. *Effects of discourse type on recall* (Prose Learning Series Research Report No. 6). Tempe, Arizona: Arizona State University, Summer 1979.
2. Meyer, B., Rice, G., Knight, C., & Jessen, J. *The effects of different discourse types on recall* (Prose Learning Series Research Report No. 7). Tempe, Arizona: Arizona State University, Summer 1979.
3. Meyer, B. *Structures of prose: Implications for teachers of reading* (Prose Learning Series Research Report No. 3). Tempe, Arizona: Arizona State University, Spring 1979.

REFERENCES

Armbruster, B. *An investigation of the effectiveness of "mapping" text as a studying strategy for middle school students*. Unpublished doctoral dissertation, University of Illinois, 1979.

Bartlett, B. *Top-level structure as an organizational strategy for recall of classroom text*. Unpublished doctoral dissertation, Arizona State University, 1978.

Britton, J., Burgess, T., Martin, M., McLeod, A., & Rosen, H. *The Development of Writing Abilities*. London: Macmillan Education, 1975.

Britton, B., Glynn, S., Meyer, B., & Penland, M. Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology*, 1982, *74*, 51–61.

Fry, E. A readability formula that saves time. *Journal of Reading*, 1969, *11*, 513–515.

McDonald, G. *The effects of instrucion in the use of an abstract structural schema as an aid to comprehension and recall of written discourse*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, 1978.

McGee, L. *Good and poor readers' ability to distinguish among and recall ideas on different levels of importance*. Paper presented at the National Reading Conference, San Diego, CA, December, 1980.

McGee, L., & Niles, J. *Sensitivity to and knowledge of text structure: Effects on children's recall of expository text*. Paper presented at the American Educational Research Association, Los Angeles, CA, April, 1981.

Meyer, B. *Organization of prose and its effects on recall*. Amsterdam: North Holland Publishing Company, 1975.

Meyer, B., Brandt, D., & Bluth, G. Use of top-level structure in text: key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, 1980, *16*, 72–103.

Taylor, B. *A comparison of good and poor readers' ability to read for meaning*. Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, 1978.

Taylor, B. Children's memory for expository text after reading. *Reading Research Quarterly*, 1980, *15*, 399–411.

Vygotsky, L. *Thought and language* (E. Hanfmann & G. Vaker, Eds. and Trans). Cambridge, MA: MIT Press, 1962 (originally published 1934).

DAVID M. MEMORY
Indiana State University

WRITTEN QUESTIONS AS READING AIDS IN THE MIDDLE GRADES: A REVIEW OF RESEARCH

It is widely believed that written questions given to readers before, during, or after the reading of a selection have an effect on how those individuals read the selections and, therefore, how well they learn the information in the selections. Though several extensive reviews of research on such questions have been published (Anderson & Biddle, 1975; Andre, 1979; Carver, 1972; Duchastel & Merrill, 1973; Faw & Waller, 1976; Hartley & Davies, 1976; Ladas, 1973), none of them has included a substantial number of studies involving middle grade students as subjects. For example, the most extensive review of this type, the one done by Anderson and Biddle (1975), referred to only three of the more than 40 reported studies of such questions used with middle grade students. Most of the investigations examined for the Anderson and Biddle review and for other reviews have involved high school and college students and adults.

The effects of adjunct questions on the reading performance of mature readers in school and out are certainly important, but the more complex and possibly more significant issues involve students who, though they have mastered word recognition skills, still have room to grow as comprehenders. It is in the middle grades that a large percentage of these readers can be found. It is then when average and superior readers can most easily and successfully be helped to develop effective comprehension strategies. Moreover, those are years when many poor learners still have a strong desire to improve as readers and can be motivated and guided by instructional techniques that help them understand what they read. Therefore, it makes sense that educators be particularly careful to have middle grade students use only adjunct questions that will at least not interfere with the development of effective reading strategies. In short, the importance of comprehension improvement among students at this level is one reason for examining the full body of middle grade research on adjunct questions.

Another reason is that no one study can be assumed to provide definitive answers to even minor issues regarding such reading aids. Even the reported studies with virtually unassailable internal validity have rarely had much external validity. Most of the adjunct questions research with middle graders has been based on single selections and single sets of comprehension questions. With research of that nature, arguments for generalization to other reading materials are tenuous at best. Moreover, most of these studies have been one-shot efforts. Students have been gathered into a central location, experimented with, and then sent on their way. Again, researchers have been hard pressed to argue that conditions such as these allow for generalization to the day-to-day reading of instructional materials in the typical classroom. For these reasons, a compilation of results from a variety of experimental settings is much more likely to yield dependable conclusions than is any single study, regardless of how well that study is designed.

Review Procedure

To prepare such a review of results, this writer searched for all reported studies in which students in the fifth through eighth grades were used to examine the effects of adjunct questions on short-term and long-term retention of information contained in or inferred from selections read by those students. Altogether approximately 80 reports whose descriptions suggested that they might be relevant to the review were purchased in microfilm or were obtained in some other form for the project. Close examination of the dissertations, ERIC documents, journal articles, and other reports revealed that 41 dealt with the variables of interest. However, eight of these studies were dropped because either they gave uninterpretable information about experimental procedures or they failed to provide the mean scores needed in the analyses. In the remaining 33 studies, over 400 comparisons involving types of adjunct questions were made that were relevant to the review.

RESULTS

Table 1 below presents the main findings of this review. In most cases results are shown only when at least four comparisons are involved. For this reason, several unreplicated comparisons yielding statistically significant differences are not represented in the table even though the comparisons are referred to separately later in this review. In the table, mean differences expressed as percentages are given only when they could be computed for representative groups of comparisons.

TABLE 1
Relative Effectiveness of Various Types of Adjunct Questions Used with Middle Grade Students

Types of Adjunct Questions Compared	Number of Comparisons Favoring Each Type of Adjunct Question	Mean of Differences as Percentage of Total Score	One-Tailed Probability
Repeated Literal Questions in an Immediate Test			
LIB vs LIA[a]	4–7	(−) 1.74	.274
LIB vs NQ	8–2	5.59	.055
LIA vs NQ	15–2	8.37	.001
LMB vs NQ	3–2	2.61	.500
LIB vs LMB	3–1	3.67	.313
LIA vs LMB	3–1	5.90	.313
LI vs NQ	7–1	7.20	.035
New Literal Questions in an Immediate Test			
LIB vs LIA	4–6	(−) 1.69	.377
LIB vs NQ	5–4	(−) .32	.500
LIA vs NQ	10–7	1.09	.314
LIA vs HIA	3–1	2.91	.313
HIA vs NQ	2–2	1.59	.688
LMB vs NQ	3–2	.60	.500
IB vs IA	1–3	.86	.313
IA vs NQ	2–2		.688
LI vs NQ	4–3	2.06	.500

(TABLE 1, *continued*)

Repeated Inferential Questions in an Immediate Test			
HIB vs HIA	2–1	4.14	.500
HIB vs NQ	2–1	3.07	.500
HIA vs NQ	3–1	(–) .74	.313
IA vs NQ	4–0	4.18	.063
New Inferential Questions in an Immediate Test			
LIB vs NQ	2–1	6.07	.500
LIA vs NQ	2–1		.500
HIB vs HIA	3–0	2.16	.125
HIB vs NQ	2–1	.96	.500
HIA vs NQ	2–2	(–) .01	.688
IB vs IA	1–2	.63	.500
IA vs NQ	3–1		.313
Repeated Literal Questions in a Delayed Test			
LIB vs LIA	1–3	(–) 2.59	.313
LIB vs NQ	4–1	3.04	.187
LIA vs NQ	3–2	5.71	.500
LI vs NQ	3–1	2.72	.313
New Literal Questions in a Delayed Test			
LIB vs LIA	2–3	.17	.500
LIB vs NQ	3–2	2.83	.500
LIA vs NQ	3–3	1.55	.656
LI vs NQ	2–2	1.75	.688
Repeated Mixed Questions in an Immediate Test			
MB vs NQ	0–5	(–) 3.11	.031
New Mixed Questions in an Immediate Test			
MB vs NQ	3–1	.89	.313
New Mixed Questions in a Delayed Test			
MB vs NQ	2–1	.45	.500

[a]Abbreviations:
L = Literal Questions
H = Higher-Level Questions
NQ = No Questions
I = Interspersed
M = Massed
B = Before
A = After

The table is divided into sections with headings which describe the dependent variable used in the comparisons of each section. For example, in the first section of comparisons, the dependent measure was literal questions that had earlier appeared as adjunct questions and that were repeated on an immediate posttest. The dependent measure in the second section of the table was literal questions that appeared for the first time as items on an immediate follow-up test. The characteristics of the adjunct questions compared are explained by the abbreviations presented at the end of the table. For example, LIA questions are literal questions interspersed after the passage sections to which they relate. HMB questions are higher-level questions massed before the whole passage.

When studying the table, it should be remembered that few of the individual comparisons represented in the data actually involved statistically significant differences. Also, the large number of binomial probability levels reported complicates the interpretation of any single result. For example, when 38 binomial tests are used, two probability levels of .05 or lower could be expected to occur by chance alone.

DISCUSSION

In general, the findings of this review suggest that the effectiveness of adjunct questions used as reading aids with middle graders is not as great as the reported effectiveness of such questions used with high school and college students and adults.

As could be expected, literal questions interspersed before or after the passage sections to which they relate seem to enhance performance on those questions repeated in immediate posttests. However, the results on new literal posttest items in an immediate test indicate that interspersed literal adjunct questions do not enhance the learning of information unrelated to those adjunct questions. This contrasts with results from studies with mainly mature readers in which literal questions interspersed after passage sections have typically led to improved performance on new literal posttest items (Anderson & Biddle, 1975). Apparently, older readers are motivated by questions interspersed after passage sections and use their mature skills to do a better job of learning information throughout a passage. On the other hand, middle grade students may have few comprehension skills to put to use when they encounter interspersed questions. Instead, they seem to focus their limited abilities on the information that answers those questions and ignore much of the other information in a passage.

Not included in the findings reported in the table are the statistically significant results from three studies that also investigated the learning of information directly stated in passages (Hatcher, 1977; LoVano, 1980; Rauscher, 1979). All three of those studies indicated that if the learning of stated information is the objective, then literal questions interspersed after passage sections are better than interspersed higher-level questions that also focus on that information. In other words, simply calling attention to stated information enhances the learning of that information by middle graders better than having the students attempt to use that information in some cognitively complex way.

Another common belief not given support by the findings of this review is that answering higher-level adjunct questions somehow prepares students to respond to other higher-level questions presented as follow-up questions. In the table the results on inferential items presented for the first time in an immediate posttest clearly do not provide support for the use of higher-level adjunct questions. On this issue one could certainly argue that three or four conflicting comparisons do not provide a reliable assessment of any certain type of higher-level adjunct question. However, two statistically significant findings not included in the table may make the results shown there more understandable. In one study higher-level questions placed before a short passage enhanced the performance of good readers, but not poor readers, on new inferential questions (Kurchak, 1978). In another study such questions interspersed after related passage sections also helped good readers, but not poor readers, in answering new inferential questions (LoVano, 1980). These two findings suggest that higher-level adjunct questions may be more effective with good readers than with poor readers in the middle grades. Possibly less skilled middle grade readers cannot understand inferential questions well enough to take advantage of them as adjunct aids.

Another statistically significant result not seen in the table seems to indirectly support these findings involving higher-level adjunct questions. In a study in which the results for good and poor readers were combined, interspersed literal questions were found to be more effective than interspersed higher-level questions in enhancing performance on new inferential questions (Rauscher, 1979). This result provides additional evidence that adjunct questions that are likely to be understood well—in this case, literal questions—are more helpful than ones that may not be understood—that is, higher-level questions.

Finally, the results shown in the table involving mixed literal and inferential items repeated in an immediate posttest deserve mention. The obvious conclusion to be drawn from these results is that providing massed questions before a passage does not enhance performance on those same questions on a follow-up test. Several explanations of this finding seem reasonable. One is that holding a list of questions in memory while reading a passage is impossible or at least very difficult. Another reasonable explanation is that trying to remember a list of questions while reading interferes with comprehension of the passage. Probably both of these explanations are partially appropriate.

CONCLUSIONS

Unfortunately, the numbers of comparisons that have been made between most types of adjunct questions have been too small to support firm conclusions about the relative effectiveness of those questions as reading aids with middle grade students. Nevertheless, three specific conclusions seem warranted:

1. Interspersed literal questions can enhance the learning of directly stated information to which those questions refer.
2. Higher-level adjunct questions enhance learning only by students who can understand and answer those questions.
3. Adjunct questions massed before a selection rarely enhance learning of stated information to which those questions refer.

REFERENCES

Anderson, R. C., and Biddle, W. B. On asking people questions about what they are reading. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9). New York: Academic Press, 1975.

Andre, T. Does answering higher-level questions while reading facilitate productive learning? *Review of Educational Research*, 1979, *49*, 280–318.

Carver, R. P. A critical review of mathemagenic behaviors and the effect of questions upon the retention of prose materials. *Journal of Reading Behavior*, 1972, *4*, 93–119.

Duchastel, P. C., and Merrill, P. F. The effects of behavioral objectives on learning: A review of empirical studies. *Review of Educational Research*, 1973, *43*, 53–69.

Faw, H. W., and Waller, T. G. Mathemagenic behaviors and efficiency in learning from prose materials: Review, critique and recommendations. *Review of Educational Research*, 1976, *46*, 691–720.

Hatcher, C. W. The effects of question-type on reading comprehension. (Doctoral Dissertation, Purdue University, 1976). *Dissertation Abstracts International*, 1977, *37*, 4984A. (University Microfilms No. 77–1724)

Hartley, J., and Davies, I. K. Preinstructional strategies: The role of pretests, behavioral objectives, overviews and advance organizers. *Review of Educational Research*, 1976, *46*, 239–265.

Kurchak, B. J. H. The effects of using an advance organizer on various levels of comprehension in fifth grade social studies. (Doctoral Dissertation, University of Colorado, 1977). *Dissertation Abstracts International*, 1978, *38*, 7263A-7264A. (University Microfilms No. 7808909)

Ladas, H. The mathemagenic effects of factual review questions on the learning of incidental information: A critical review. *Review of Educational Research*, 1973, *43*, 71–82.

LoVano, G. M. The relationship of self concept and interspersed questions to aid comprehension in the processing of textual materials. (Doctoral Dissertation, Case Western Reserve University, 1980). *Dissertation Abstracts International*, 1980, *41*, 1494A. (University Microfilms No. 8021702)

Rauscher, W. C. An investigation into the facilitative effects of two kinds of adjunct questions on the comprehension of students at two grade levels during the reading of textual materials. (Doctoral Dissertation, University of Iowa, 1978). *Dissertation Abstracts International*, 1979, *39*, 5937A-5938A. (University Microfilms No. 7902934)

JUDY NICHOLS MITCHELL
JOHN M. BRADLEY
WILBUR S. AMES
University of Arizona

ORIGINAL VERSUS ADAPTED SHORT STORIES: THE PERCEPTIONS OF ADOLESCENT READERS

An established secondary school teaching practice is the use of adapted or rewritten text materials as opposed to their respective original versions. The assumption is that these adapted materials are less difficult and hence more appropriate for use by less able readers. Many teachers feel that adapted texts offer these students the same concepts and main ideas presented in the original materials while increasing reading ease.

However, the use of text analysis techniques to compare adapted texts with their original versions has revealed many differences such as sentence restructuring, content change, lexical change, and changes in rhetorical devices (Davison, et al., 1980). The fact that such differences exist raises the question whether adapted texts may in fact be more difficult for readers than the original versions, since less information may be provided in text adaptations and readers may therefore need to make more inferences.

One approach to the investigation of this question is the comparison of perceptions of students relative to difficulty and preference after reading both an original and an adapted text. This study was undertaken to investigate whether adolescent readers, in general, are 1) able to perceive text difficulty differences between original and adapted versions of the same story, and 2) prefer reading adapted versions to original story versions.

METHOD

Subjects

Seventh and eighth grade students from two schools in Tucson, Arizona (N = 522) were selected to participate in the study. The school populations were predominantly Anglo in one school and Hispanic in the other.

Materials

Two short stories and their three published adaptations were selected to control for story-specific findings. The stories selected were "Holiday Man," by Richard Mathewson and "Hearts and Hands," by O. Henry, the latter having two published adaptations. While the content and setting of the stories are different, both are characterized by surprise endings. According to the Fry formula (1977), the readability level of both versions of "Holiday Man" (HMO and HMA) is second grade; that for the original version of "Hearts and Hands" (HHO) is sixth grade, the first adaptation (HHA1) second grade and the second adaptation (HHA2) fourth grade. These stories were judged appropriate for use in the study since the original versions of both have been included in several anthologies designed for junior high readers and the adapted versions have appeared in published high interest-low vocabulary materials (HMA in Allyn and Bacon's *Breakthrough* series, HHA1 in Scholastic *Action*, HHA2 in Scott Foresman *Tactics*).

Procedures

During one class period of 45 to 50 minutes, subjects were asked to read silently an original and an adapted version of the same story. Following their reading, subjects responded to a rating inventory containing three pairs of 9-point Likert-type ratings designed to address student perceptions of text difficulty, story interest and story preference for each of the two story versions.

Story sets were randomly distributed and order of original/adapted presentation was counterbalanced, resulting in six treatment groups of 87 subjects each. Therefore, each story adaptation was read with its original version in both possible orders.

RESULTS

Originally, it was intended to collapse the treatment groups by story version to analyze mean responses for each of the three pairs of Likert-type items. Each story version had been read first by one treatment group of 87 subjects and second by another treatment group of 87 subjects. The mean ratings for difficulty, preference and interest for the six treatment groups for all story versions are presented in Table 1.

However, when mean rating data were computed for all items, it was apparent that there was an order effect. Therefore, two-way ANOVA procedures were conducted by the original versions only and then by the adapted versions only.

Original/Adapted Analysis

For both the preference and the interest items and for both original and adapted versions, the second story read received significantly higher ratings, regardless of version. For the difficulty items, five or six story versions were rated significantly easier when read second than when read first. In addition, mean difficulty ratings reflected actual story version readability differences according to the Fry formula. Specifically, HMO and HHA1 were rated as significantly less difficult than HHO and HHA2, respectively. All reports of significance were at the .01 level.

TABLE 1
Mean Ratings for Difficulty, Preference and Interest of Original and Adapted Story Versions by Six Treatment Groups (N=87 per group)

Treatment		Difficulty		Preference		Interest	
First Version Read	Second Version Read	First Version Read	Second Version Read	First Version Read	Second Version Read	First Version Read	Second Version Read
HMO	HMA	3.61	2.52	5.48	5.53	4.87	5.08
HMA	HMO	2.74	2.87	4.21	6.11	3.68	5.72
HHO	HHA1	4.61	2.74	5.09	6.34	4.37	5.85
HHA1	HHO	2.71	3.70	5.52	5.95	5.07	5.49
HHO	HHA2	4.32	2.88	5.14	5.51	4.45	4.98
HHA2	HHO	3.47	3.55	4.93	6.13	4.09	5.23

First and Second Readings Analyses

Because of the strong order effect discussed in the previous section, separate analyses of variance were conducted by first and second readings. For both readings, original versions were rated as significantly more difficult than their respective adaptations. Further, a strong story effect was observed for both readings for HHA1, which was rated significantly higher on both the interest and preference dimensions than all other story versions.

CONCLUSIONS

A significant order effect was observed across all six items for all six treatment groups. Subjects rated all items differently depending on whether they had read a given story version first or second. Specifically, subjects favored the last story read in rating how much liked or how interesting they found the original and adapted versions, regardless of which version was read last. Therefore, no overall preferences were evident for adaptations over original versions, contrary to expectations of teachers of participating subjects.

Further, although subjects were consistent in rating all the original story versions as more difficult than any of the adaptations, order also affected subjects' ratings of the relative difficulty level of the versions they read. Specifically, when an original story was read first, that version was rated as much more difficult than its adapted version. For two of the remaining conditions, when the adaptation preceded the original, the latter was perceived as only slightly more difficult than its adaptation. Therefore, subjects do perceive which of two story versions is easier, regardless of the order in which they are read, but their ratings are strongly affected by the order of presentation of the versions.

Story effects attributable to individual differences in story versions were observed for HHA1. On the items addressing the preference and interest dimensions, if the original version was read first, the adapted versions were rated only slightly higher in the case of HMA and HHA2, but considerably higher for HHA1. Conversely, if the adaptation was presented first, the original version was rated much higher for HMA and HHA2 but only slightly higher for HHA1. Clearly, although the main order effect trend was observed, subjects were responding differently to the combination of HHO and HHA1 than they did to the other conditions, regardless of order of presentation.

IMPLICATIONS

The following implications appear to be related to the above conclusions:

1. Since students rate the difficulty level of original stories differently according to the order of presentation, perceived difficulty may be related to the degree of familiarity students have with the story. Therefore, a question remains as to whether this familiarity can be gained only by reading an adaptation followed by the original version, or whether two readings of the original would produce the same effect. A further study is planned to investigate the effects of these two treatments.

2. Since interest and preference are affected by order, they appear to be relative rather than absolute values, not always associated with text difficulty. Therefore, teachers may wish to select class materials at the secondary level according to criteria other than difficulty level.

3. Student ratings of difficulty, interest and preference vary according to the story presented. Therefore, researchers should utilize more than one set of passages or treatment in their designs. Had this study been based solely on a comparison of HHO and HHA1, the results and implications would be quite different and far less generalizable.

4. The same trends were observed regarding difficulty ratings of an original/adapted pair of the same readability level (e.g., HMO and HMA) and those of different readability levels (e.g., HHO and HHA2). Therefore, it appears that readers may be responding to text characteristics other than those identified by readability formulas. Further research needs to be conducted in this area.

5. Finally, it is curious that an artifact of the design of the study, that of counterbalancing treatment groups for order of presentation, actually resulted in the study's main effect. Implications for researchers are that when you cast your net upon the waters, you may catch some types of fish that you didn't anticipate.

REFERENCES

Davison, A., Kantor, R. N., Hannah, J., Hermon, G., Lutz, R., and Salzillo, R. *Limitations of readability formulas in guiding adaptations of texts*. (Tech. Rep. No. 162) Urbana: University of Illinois, Center for the Study of Reading, March, 1980.

Fry, E. B. Fry's readability graph: clarifications, validity, and extension to level 17. *Journal of Reading*, 1977, *21*, 242–252.

ALLAN R. NEILSEN
BARBARA J. RENNIE
ARLENE M. CONNELL
University of Calgary

ALLOCATION OF INSTRUCTIONAL TIME TO READING COMPREHENSION AND STUDY SKILLS IN INTERMEDIATE GRADE SOCIAL STUDIES CLASSROOMS

The purpose of this study was to determine the proportion and distribution of time devoted to reading comprehension and study skills during social studies instruction. Durkin (1978–79) found that 12.81% of fourth grade social studies instructional time was spent on comprehension-related activities. However, no instances of direct comprehension instruction were observed. A similar pattern emerged for study skills; one percent of class time was devoted to study skills-related activities but, again, no instances of direct instruction were observed. These findings suggest that fourth grade social studies teachers assume that students have already learned to read and that they are able to apply their reading and study strategies to derive meaning from their texts—with no additional instruction or provision for transfer of strategies being necessary. The present study attempted to replicate the findings for a subset of the instructional practices reported by Durkin in order to assess the reliability and generalizability of her results.

METHOD

Subjects

A total of eight social studies teachers and sixteen students from five schools participated in the study. All teachers were volunteers and represented grades four (N=2), five (N=4), four-five split (N=1) and six (N=1). Four of the schools are located in a major metropolitan centre while the fifth is located in a satellite community. The areas in which the schools are located can be characterized low-middle to high-middle socioeconomically.

Procedures

A modified subset of Durkin's (1978–79) categories of comprehension and study skills instruction provided the basis for classifying teachers' classroom behaviours. Only twenty categories were used in the present study and those were restricted to teacher behaviours. These restrictions on the breadth and depth of the study were imposed for two reasons. First, the proportion and relative distribution of class time devoted to comprehension and study skills (i.e. teacher behaviour) was of primary interest. Therefore, student behaviour was not included. Second, and ultimately more important, the number of categories was reduced to facilitate accurate, consistent recording.

In addition to limiting the number of categories, some of Durkin's original categories were reorganized in order to help clarify the focus/emphasis of the observations for purposes of interpretation. The major change involved the creation of a category called Not Text-Based which included all activities involved in the presentation of subject-specific information by means other than written text (e.g., lectures, discussions). It also subsumed all of Durkin's original Assignment-related behaviours (i.e., "checks", "gives", "helps with"). This category was created in an attempt to characterize overall instructional behaviour more accurately. An exclusive focus on Text-Based comprehension and study skills-related behaviours can lead to the inappropriate conclusion that teachers are spending very little of their class time constructively.

Another point of departure from Durkin was the method used to record observations. Whereas Durkin and her colleagues wrote down all observed behaviours and then presumably classified them in a post hoc manner, the observers in this study used a checklist with all categories listed on one page. Thus a decision regarding a behaviour was made at the time of actual observation. The checklist information was supplemented with descriptive comments about what was being observed. All information was recorded for blocks of 15 seconds rather than 30 seconds since it was felt that the shorter unit would be much more sensitive to the often rapid changes in classroom activity.

Data Collection

Each of the eight social studies teachers was observed during three successive lessons (i.e., 24 lessons; 973 minutes). The amount of time spent on reading comprehension and study skills instruction was recorded for each lesson and a cumulative summary completed for each teacher as well as the total group. Most lessons (i.e., 19 lessons; 762 minutes) were co-observed. Logistical problems in the observers' schedules accounted for the five lessons that were not co-observed. Following each co-observed session, the observers met to compare their records and notes and to resolve any discrepancies. Inter-observer reliability was .94.

RESULTS

Text-Based Instruction

Comprehension An average of 22.28% of teachers' class time was devoted to comprehension-related activities. Most (15.03%) of this time was "Comprehension: helps with assignment". It is important to note that the amount of time spent helping students with text-based assignments varied considerably across teachers (from 66.67% for Teacher 5 to Not Observed for Teachers 2 and 3 who did not use text materials in any of the classes observed). A further caution in interpreting the data for "Comprehension: helps with assignment" is necessary. It is possible that the findings for this category are inflated since it was not always possible to determine exactly what the teacher was saying to students (i.e., the observers' positions in the classroom were fixed while the teacher moved around to individuals and groups). If students were working on a text-based assignment and receiving help from the teacher, the behaviour was recorded as "Comprehension: helps with assignment" even though some of this time might have involved comprehension instruction, assessment, formulating predictions or other types of "help".

The only other comprehension-related behaviour receiving much time was "Comprehension: assessment". Although it represented an average of 5.02% of teachers' class time, pronounced individual differences were evident. No assessment was observed for Teachers 1 and 2 whose instruction was not text-based or for Teacher 5 who did not explicitly question students about their understanding of text but instead had them spend most of their time working independently in pairs on a text-based assignment. All assessment observed involved questions that were assigned orally, written on the board or were assigned from texts.

Virtually no time (.64%) was spent on direct instruction of comprehension strategies or on preparing students for reading assignments (1.08%). Teacher 8 was the only one to devote any time (4.48%) to direct instruction. Further, Teachers 6, 7 and 8 were the only ones who devoted any time (3.86%–1.12%) to preparing students for reading assignments. Since this study was conducted relatively late in the school year, it is possible that any direct instruction related to comprehension strategies had been conducted previously. Partial support for this explanation was provided by Teacher 5 who indicated that she had spent a lot of time in September teaching students how to write summaries. Another important consideration to be made when interpreting these findings is the fact that text materials were used in only 13 of the 24 lessons observed, thus limiting the opportunity for text-based comprehension and study skills-related activities to take place.

Study Skills Although most teachers spent some time on study skills-related activities, the total time spent varied greatly across teachers (9.32%–0%). Also, no discernible emphasis emerged across the sub-categories observed. Perhaps the most consistent attention was given to the application of study skills which, in this study, usually involved the use of an atlas or almanac rather than strategies associated with expository prose (e.g., skimming, outlining, note-taking).

Not Text-Based Instruction

Behaviour that was not text-based accounted for an average of 35.25% of teachers' class time with seven out of the eight teachers involved to a greater (61.18%) or lesser (17.36%) extent. The exception, Teacher 5, had her students involved with text during all lessons observed.

By far the majority of time in this category (31.78%) involved activities classified as "Not text-based: instruction". It should be noted that this category included more than instruction per se; it subsumed virtually all of the same subcategories as Comprehension Skills (i.e., "application", "assessment", "assignment", etc.) with the exception of "helps with assignment". An average of 3.47% of teacher time was classified as "Not text-based: helps with assignment".

DISCUSSION

Findings of the present study for the amount of time devoted to "Comprehension: instruction" and "Study Skills: instruction" are virtually the same as Durkin's. In both studies virtually no direct instruction in these areas was observed.

Adopting the broader definition of comprehension instruction suggested by Hodges (1980), the present study found that almost twice as much time (i.e. 22.28%) was devoted to comprehension-related activities as a whole as that reported by Durkin (i.e., 12.82%). However, most (13.25%) of this difference can be accounted for by the greater amount of time devoted to helping students with assignments in the present study. Study skills-related activities accounted for 3.19% of instructional time in the present study compared to 1% reported by Durkin.

Although it is not possible to make any strong comparisons of findings from the two studies due to differences in sample size and total time observed (one-third as much for both in the present study) and differences in procedures, it can be argued that the results are similar (and for direct instruction, almost identical). However, the really important question is whether these findings represent an acceptable or even reasonable level of performance on the part of teachers. Durkin's (1978–79) discussion leaves the strong impression that teacher input in content areas is inadequate. Although this is an intuitively persuasive conclusion, it does not take opportunity or appropriateness of instruction into account.

The fact that Not Text-Based instruction accounted for 35.25% of instruction observed in the present study suggests that it is a significant instructional option exercised by teachers. As a result, the opportunity for text-based instruction to take place is reduced. The "low" proportion of time observed for direct comprehension and study skills might be, in part, a reflection of the amount of time spent on Not Text-Based instruction.

Further, even when instruction *is* Text-Based (i.e., 54.17% in this study), it might not be necessary or even appropriate for direct comprehension or study skills instruction to take place. After each text-based observation session in the present study, the observers made a judgment regarding the appropriateness of instruction. Two considerations were involved in judging appropriateness. First, was Comprehension and Study Skills instruction/preparation necessary for understanding the text? If not provided when deemed not necessary, instruction was judged to be appropriate. Second, if deemed necessary, was sufficient help provided? If instruction/preparation was provided but was not extensive enough or not directly relevant to the immediate needs of the students, it was judged to be inappropriate. In only three (i.e., 23.08%) of the 13 lessons in which instruction was text-based was "Comprehension: Instruction" observed. While two (i.e., 15.39%) of the remaining text-based lessons *were* judged to need direct instruction/preparation, eight (i.e., 61.54%) were judged to be appropriate without direct intervention. Thus, it seems reasonable to suggest that the proportion of Not Text-Based instruction and the appropriateness of instruction that *is* Text-Based be carefully considered when attempting to determine whether the amount of instructional time devoted to comprehension and study skills in social studies and other content areas is adequate.

REFERENCES

Durkin, D. What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 1978–1979, *14*, 481–533.

Hodges, C. A. Toward a broader definition of comprehension instruction. *Reading Research Quarterly*, 1980, *15*, 299–306.

M. C. WITTROCK
University of California, Los Angeles

THREE STUDIES OF GENERATIVE READING COMPREHENSION

Introduction

To read a text with understanding, I maintain that one must actively generate relations (1) among the parts of the text, and (2) between the reader's knowledge or experience and the text. The written message provides the instructions that guide the reader's thought processes toward the construction of these two types of relations. The reading teacher complements the text by individualizing and leading these generative activities, and by teaching children how and when to perform them. Even ideas, concepts, principles, and answers given in the text to the readers must still be generated by them for comprehension to occur. To understand them, the readers must construct relations among the parts of the text and between these parts and their knowledge or experience.

Generative Teaching

In this paper I want to focus on the data of three recently completed studies on the teaching of generative reading to school children. Before discussing these studies on the teaching of children to construct summaries, headings, and analogies involving their experience, I would like to state a couple of assumptions we made about motivation and attention, which we did not explicitly vary in these studies.

Because generative learning occurs when learners actively construct relations, as described earlier, certain types of motivation or attribution are important for facilitating these activities. The learners should believe that through their effort at constructing relations comprehension will occur, or at least will be facilitated. They should attribute success or failure in understanding the text, at least in part, to their own effort. They should assume some responsibility for reading with comprehension. They should believe that comprehension will not occur automatically with time or with practice.

For teachers, the same type of attribution and responsibility for teaching should exist. Through their efforts at designing and conducting generative teaching activities, they should expect reading comprehension to occur or to be facilitated. Teachers' efforts facilitate learners' comprehension when they enhance the readers' construction of relations involving the text and the readers' knowledge.

Attention, the second cognitive process we did not study, guides the learners to the construction of relevant and useful relations. Although many kinds of constructed interactive relations among stimuli facilitate memory, as occurs in mnemonic devices using imagery, the relations which facilitate comprehension involve coherent and relevant associations between knowledge and the text.

Generation, the process studied in the three experiments discussed now, is central to the model of generative teaching. Several methods have been used to stimulate it. These methods involve the text, the teachers, and the strategies developed in the learners. The goals of the studies were (1) to teach the learners how and when to construct different types of relations, and (2) to measure the results upon comprehension that these teaching procedures produced. Each of the three studies involved individual, random assignment of learners to treatments, reading materials commonly used to teach reading, and elementary or junior high school students enrolled in large, urban public schools. Time to learn was held constant across all treatments. We also recorded the number of sentences generated by the learners during instruction.

Three Studies

The data of the first study, which have been published previously (Doctorow, Wittrock, & Marks, 1978), will be briefly mentioned. In this study, children were asked to read stories (1) with or without paragraph headings, and (2) with or without instructions to generate a summary sentence for each paragraph. The data indicated sizable effects upon memory and comprehension produced by either or both of these instructional treatments. Comprehension was measured by multiple choice tests of inferential meaning, whose odd-even reliability coefficients were .93 and .96 (see Doctorow, et al., 1978 for sample items). The greatest facilitation, which approximately doubled comprehension, was produced when the junior high school students constructed summary sentences using the paragraph heads we gave them. These data support the model's predictions regarding the facilitative effects of generating summary sentences, especially when a paragraph heading, which serves as a structure to guide generation, is used in the construction.

The data of the second study have not yet been published. In this complicated study with the same materials and tests described above and several hundred junior high school students, the control group was given two stories to read without any generative activities to perform. The 12 experimental groups were given the same stories to read and either paragraph headings to guide their generations or instructions which asked them to generate topic sentences and sentences relating the text to their knowledge, or both of these instructional aids—paragraph headings and instructions to generate sentences.

Although there are a number of interesting statistically significant differences in these data, the most sizable difference among the means occurs between the control group and all or any one of the experimental groups. The generative procedures, including the paragraph headings, facilitated comprehension and retention. There was no exception to this finding. It occurred across all four tests and four stories. On the test of comprehension of the first story, the control group scored 10.4, while the 11 experimental group means ranged from 12.0 to 14.1. The corresponding means on the comprehension test over the second story were 9.0 for the control group, and 11.1 to 12.4 for the experimental groups. On the retention test over the first story, the control group mean was 8.7, and the experimental group means ranged from 11.2 to 13.0. On the retention test over the second story, the control group mean was 8.3, and the experimental group means ranged from 17.0 to 19.6. The differences among the experimental groups were usually smaller than the differences between the control group and the experimental groups. It seems that the process of generating relations between the parts of the text or between the text and experience is the most significant variable. The different methods we used to stimulate this process were all effective and their effects were not greatly different from one another.

We also collected data on individual differences in verbal and spatial cognitive processes. We are now analyzing these individual difference data to see if the treatments work differentially well with the different cognitive abilities.

The last of the three studies I want to discuss is quite different from the other two experiments, although it uses the same theory and experimental methods. It will soon be published in *Reading Research Quarterly* (Linden & Wittrock, in press). In this study, we used classroom teachers instead of written materials. We randomly assigned 58 children to the four treatments which involved three days, one hour per day, of classroom instruction. The first group of students was a control group taught by the children's regular teacher, who was free to use whatever teaching methods she chose. The second group was a control group taught by Michele Linden, an experienced reading teacher. The third and fourth groups, the two experimental groups, were also taught by Dr. Linden. In these two groups, the children constructed verbal and spatial representations for the ideas of the stories they read. One of these groups began with spatial generations and proceeded to verbal generations. The second of these two experimental groups began with verbal generations and proceeded to spatial generations.

The two generative groups averaged 28.6 and 31.3 items correct on the test of comprehension. The two control groups averaged 17.7 and 21.6 items correct. These sizable differences between the experimental and control groups are statistically significant ($p<.01$) in the direction predicted by the model of generative learning. Time to learn was held constant across all groups of learners.

As a whole, these three studies, and others we have recently done, consistently show that generative teaching activities facilitate comprehension among children in classrooms using common reading materials. Differences among our generative activities do not account for as much facilitation as does the difference between a generative activity and one that either represents a standard teaching procedure or one that is identical to the generative treatment in every way, except that it does not involve a generative activity by the learner.

Discussion

These findings support predictions of the model and imply that reading comprehension can, sometimes at least, be enhanced by using the principles of teaching built into these studies.

For testing learners in reading classes, this model of generative learning also implies some changes in procedures. To teach within this model, one needs to measure the types of generation, such as verbal or spatial, children can engage in either spontaneously or upon request. The teacher also needs to know whether children can relate parts of the text to one another or to their knowledge and experience, again, either spontaneously, or upon request.

These generative abilities could form a developmental progression, beginning with the youngest school children and their need for directions to stimulate generation to the ability of older students to generate spatial and verbal, superordinate, coordinate, and subordinate relations either spontaneously or upon request. Tests of generative abilities would look different from some of today's diagnostic tests. They could play an important role in the design of generative teaching, and in the teaching of children to use strategies of generation to facilitate comprehension.

ACKNOWLEDGEMENTS

This paper was reported at a symposium that was funded, in part, by a California Policy Seminar grant. The grant is to Harry Singer (University of California, Riverside), John McNeil, and Merl Wittrock (University of California, Los Angeles), and Robert Ruddell (University of California, Berkeley) to study "Achievement in Testing Basic Skills."

REFERENCES

Doctorow, M. J., Wittrock, M. C., & Marks, C. B. Generative processes in reading comprehension. *Journal of Educational Psychology*, 1978, *70*, 109–118.

Wittrock, M. C. Learning as a generative process. *Educational Psychologist*, 1974, *11*, 87–95.

Wittrock, M. C. The cognitive movement in instruction. *Educational Psychologist*, 1978, *13*, 15–30. (a)

Wittrock, M. C. Reading Comprehension. In F. J. Pirozzolo & M. C. Wittrock (Eds.). *Neuropsychological and cognitive process in reading*. New York: Academic Press, 1981.

Wittrock, M. C., & Carter, J. Generative processing of hierarchically organized words. *American Journal of Psychology*, 1975, *88*, 489–501.

Wittrock, M. C., & Lumsdaine, A. A. Instructional psychology. *Annual Review of Psychology*, 1977, *28*, 417–459.

Wittrock, M. C., Marks, C. B., & Doctorow, M. J. Reading as a generative process. *Journal of Educational Psychology*, 1975, *67*, 484–489.

BONNIE A. CARROLL
PRISCILLA A. DRUM
University of California,
Santa Barbara

THE EFFECTS OF CONTEXT CLUE TYPE AND VARIATIONS IN CONTENT ON THE COMPREHENSION OF UNKNOWN WORDS

The present paper attempts to examine the effectiveness of two different categories of context clues embedded in three different types of text in facilitating word comprehension by readers of different ability levels. The importance of context as a determinant of word meaning has been emphasized by researchers in the field of reading who have stressed the relationship between the use of context and vocabulary development (Dulin, 1969). Several different kinds of studies have explored the nature and use of context clues. A majority of studies have been based on a generalized response to context rather than a specific response to particular kinds of context; overall conclusions drawn are that context clues do facilitate the comprehension of word meaning, that there is a high positive correlation between ability and effective usage, and that instruction in clue recognition and usage is significant in effective clue usage (Dulin, 1969; Biemiller, 1970; Weber, 1970; Gipe, 1978; Carroll, Madison, and Drum, 1980).

A second type of study has investigated the effective placement of target words within sentence context for maximum clue value. Aborn and Rubenstein (1957) discovered that context is generally most effective when the unknown target word is placed in the center of a sentence. Madison, Carroll, and Drum (1981) studied instead, the placement of clues relative to the target word, looking specifically at directionality (clue before and after the target word) and proximity (relative closeness to the target word). They concluded that direction influenced clue usage, being more effective in the after position for high ability students and in the before position for low ability students, and that close proximity to the target word made the clue more effective for both ability groups.

A third type of study has explored the area of classifying clue types. Ames (1966) found 14 unique categories of clues by post hoc analysis of student responses. His graduate student subjects read fiction stories and non-fiction articles with pseudowords substituted for every 50th word, each containing the same number of letters as the original word and composed of syllable combinations that approximated a true English word. The subjects read each passage and then responded in personal interviews as to what the real English word should be and their reasons for it. Ames, however, did not make an attempt to analyze the effectiveness of clue type nor the effectiveness of text type in which the clues were embedded.

Quealy (1969) replicated Ames' study with senior high school students and concluded that his subjects used the same clues, although with less accuracy. Rankin and Overholser (1969) also replicated Ames' study with upper elementary students but used a cloze technique rather than substituting nonsense words and an objective test to measure clue sensitivity. They conducted a more complete analysis of subjects' use of clues and found a difference in the frequency in which clue types were used. In none of the context clue studies do the authors consider the overall effect of passage type in which context clues are embedded.

This study was designed to examine the effective use of the two most frequently mentioned types of context clues—definition and synonym—(as determined by Rankin and Overholser), embedded in three different types of text—fiction, biography, and scientific information—in determining the meanings of unknown words by 8th grade students. Also, the directionality variable, clue before or after the target word, was reexamined for ability effects. It was predicted that type of text would influence clue usage. Since scientific information passages are generally more complex and detailed, the difficulty of the text would interfere with the use of clues, so that fiction, biography, and science would rank from more to less effective. It was also predicted that clue type would influence the use of clues, with definition being more influential than synonym, based on Rankin and Overholser's findings. It was expected that higher ability subjects would make more effective use of clues than lower ability subjects, but that there would be an interaction between ability and directionality of clue, as reported in Madison et al. (1981).

METHOD

Subjects

Subjects participating in this study were 24 8th grade students from a junior high school of lower-middle to middle SES. They were divided into four groups by ability and sex (i.e., group 1—high ability females, group 2—high ability males, group 3—low ability females, and group 4—low ability males). Level of ability was determined by teacher evaluation and corroborated by standardized test score results (California Test of Basic Skills).

Materials

Materials consisted of three passages, each passage approximately 200 words in length, and varied as to topic, author, and style. Passage 1 was a fiction, boy-meets-girl story; passage 2 was a biography of a famous baseball player; and passage 3 contained scientific information, a description of the planet Earth. Four content words were deleted from each passage, and a pseudoword composed of syllable combinations that would approximate a true English word was substituted. Context clues for these pseudowords were systematically varied by type and direction. Two types of clues, definition and synonym, were evenly divided over the target words, with half the clues coming before the word and half after the word. Proximity was controlled; all the clues were within five words of the target pseudoword. All three passages were taken from 5th grade level curriculum materials. Passages were balanced across trials to control for position bias. Two scorers reviewed and agreed on the appropriateness of the clue type and content type.

Procedure

Following Ames' and Quealy's studies, an introspective approach was used to collect data; each student was interviewed separately, with no time limit imposed, and their answers were tape-recorded. After being instructed that they would read three passages with four nonsense words substituted in each and asked to supply a true English word for each pseudoword and the reason for the replacement, each student read each passage silently and then responded as to 1) what English word would fit the place of the nonsense word, and 2) what specific clue within the passage helped to determine what the real word should be. They were given the opportunity of making notes if they wished on the written passages to aid in their answering.

In evaluating the word replacement responses, the exact word or acceptable dictionary synonym with correct tense and grammatical structure was required. In evaluating the clue identification responses, the specific clue embedded in the passage was required. The scoring procedure was a 0 for a wrong answer and a 1 for a right answer for both word replacement and clue identification.

RESULTS AND DISCUSSION

The research design included one between subject variable, ability (2 levels), by text topic (fiction, biography, and scientific information), with directionality (before and after) and clue type (definition and synonym) crossed with passages. There were two dependent measures, word replacement and clue identification. A multivariate analysis was used to examine the data. This section will be examined first for ability, followed by text topic, clue type, and directionality.

Ability

As expected, high ability students outperformed low ability students, *F* $(1,20)=28.1$ $p<.00$. Differences between ability groups was demonstrated during the interviews in several ways. Time required for completion of the task varied between groups, with the average time per interview for high ability students being 12 minutes and 20 minutes for low ability students. Strategies used in responding also differed. Low ability students at times used the pseudoword itself as a clue to the real word, choosing a look-alike rather than a word to fit the context; for instance, the *fomme* pseudoword provided the word *form* rather than the expected *layer* or a synonym. There was also a tendency for low ability students to use as clues words immediately around the target word without regard to the context of the passage; for example, "Even though Lou was a *heag* player . . . " brought responses such as *basketball* when the correct word was *poor*. Contrary to Madison et al., there was no interaction between ability and clue direction, as the before position was more effective for all subjects. Correlation between clue identification and word replacement scores was .77, showing that if the reader did well on one, he also did well on the other.

Text Topic

Type of text was highly influential, *F* $(2,40)=4.2$, $p<.02$, with fiction and the biography passages receiving more correct responses than the passage containing scientific information. This may be due to the precision of concepts in science, requiring an exact answer, while many concepts may fit fiction or a biography. An interaction between text topic and dependent measures, *F* $(2,40)=4.03$, $p<.03$, showed that for the fiction and biography passages, the clue identification scores were higher, while the word replacement scores were higher for the scientific passage. (See Table 1.)

Clue Type

Clue type did not substantially differ overall. The fact that clue type did not make a difference may be due to the relative similarity of definition and synonym type clues. More control in selecting contrasting clue types might pro-

TABLE 1
Averages for Text Type by Dependent Measures, Clue Types, and Clue Direction

	Fiction		Exposition		Science	
	M*	SD	M*	SD	M*	SD
Dependent Measures						
Word	2.92	0.82	3.13	1.41	2.63	0.89
Clue	3.08	0.82	3.29	1.63	2.00	0.98
Clue Type						
Synonym	2.58	1.33	3.38	1.33	2.96	1.17
Definition	3.42	1.03	3.04	1.83	2.54	1.52
Clue Direction						
Before	3.46	1.03	3.54	1.97	2.42	1.33
After	2.54	1.26	2.88	1.60	3.10	0.98

*out of a possible 4

duce different results. A further consideration is the difficulty in completely controlling passage clues while attempting to embed specific clues; the context of the passage itself provides clues. There was an interaction between passages and clue types, F (2,40) = 5.67, $p<.01$, showing that for the fiction passage, the definition clue was more effective for a correct response, while in the exposition and science passages, the synonym clue was more effective (see Table 1).

Clue Direction

Direction of the clue relative to the target word made a difference overall, F (1,20) = 5.33, $p<.03$, with clues in the before position being more effective, as previously mentioned. An interaction between text topic and clue direction, F (2,40) = 8.79, $p<.00$, showed that the before clue was more effective for both the fiction and biography passages, while the after clue was more effective for the science passage (see Table 1). A reason for this may be that the position of the clue lends itself to reader prediction of the content. Both the fiction and biography passages were less detailed than the scientific passage and likely made fewer demands on the reader's attention; the reader could anticipate the upcoming text. The science passage, on the other hand, is more precise in its wording, expecting a definite answer. Until the reader comes to the nonsense word, attention is focused on the details of the text; at the nonsense word, the reader may then begin to look in the text to explain it.

EDUCATIONAL IMPLICATIONS

Major findings of this study showed that high ability students were consistently better in selecting the correct clue and responding with the correct response or a synonym than were lower ability students. Text content also proved influential, as the fiction and biography passages received more correct responses than did the passage containing scientific information. While clue type did not differ overall, direction of the clue did make a difference, with the before position clues being more effective than the after position clues.

Empirical evidence has demonstrated the effectiveness of context in facilitating word comprehension; the ability to use context well, in fact, underlies mature reading. While students are encouraged to use the context of a reading selection, their efforts in being sucessful may vary according to their mastery of the skill; instructional implications would be, then, for the direct teaching of clue recognition and employment, together with appropriate practice. In instruction, ability level of the student should be considered, as this study has demonstrated the less effective use of clues by low ability students, even with appropriate level reading materials. Developing meaning through using contextual aids has been described as a problem-solving task, which is correlated highly with high ability students; low ability students, then, require more instruction and practice. Text topic should also be considered with attention as to reading difficulty and text purposes, e.g., fiction for casual reading as compared to technical for specific details; knowing what, where, and when to look for clues is based on the particular selection being read. While the use of context will not necessarily provide specific and multiple meanings of a word, the student should recognize that context surrounding a textual segment can help in determining the meaning of that segment.

REFERENCES

Aborn, M., & Rubenstein, H. Word class distribution in sentences of fixed length. *Language*, 1957, *32*, 666–674.

Ames, W. S. The development of a classification scheme of contextual aids. *Reading Research Quarterly*, 1966, *2*, 57–82.

Biemiller, A. The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly*, 1970, *6*, 75–96.

Carroll, B., Madison, J., & Drum, P. *The effect of context on the comprehension of unknown words*. Unpublished study, 1980.

Dulin, K. L. Using context clues in word recognition and comprehension. *The Reading Teacher*, 1969, *23*, 440–445, 469.

Gipe, J. P. Investigating techniques for teaching word meanings. *Reading Research Quarterly*, 1978, *14*, 624–644.

Madison, J., Carroll, B., & Drum, P. The effect of directionality and proximity of context clues on the comprehension of unknown words. *National Reading Conference Yearbook*, 1982, this journal.

Quealy, R. J. Senior high school students use of context aids in reading. *Reading Research Quarterly*, 1969, *4*, 512–533.

Ranking, E. F., & Overholser, B. M. Reactions of intermediate grade children to contextual clues. *Journal of Reading Behavior*, 1969, *1*, 50–73.

Weber, R. M. A linguistic analysis of first-grade reading errors. *Reading Research Quarterly*, 1970, *5*, 427–451.

JOSEPH L. VAUGHAN
East Texas State University

GAYLE CASTLE
Southwest Oklahoma State University

KAREN GILBERT
Arlington (Tx) Indep. School District

MARY LOVE
Garland (Tx) Indep. School District

VARIED APPROACHES TO PRETEACHING VOCABULARY

The relationship between vocabulary knowledge and reading comprehension is as "common-sense" obvious as it is empirically verified. From a pedagogical view, however, few issues are as enigmatic. While it is clear that vocabulary is a semantic conveyance, efforts to enhance readers' comprehension through vocabulary instruction have often gone for naught (Anderson and Freebody, 1979; Tierney and Cunningham, 1980; Ahlfors, 1980).

Several factors seem predominant within this enigma. First is the concept of "word" and what it means to know a word. Vygotsky (1978) and Donaldson (1978) emphasize that word meanings are dynamic and fluid, not static. Words "mean" differently to different people and words "mean" differently in different contexts. Estes and Vaughan (in press) note that words do not inherently have meaning; instead people have meanings for words. A second element that seems to confound the vocabulary/comprehension enigma is the words chosen for instruction. Graves (1978) has delineated four potential relationships between a word (and its meaning) and a learner. This relationship could/should serve as the primary determiner of what words should be taught; unfortunately most vocabulary studies seem to have been conducted (and most vocabulary lessons?) with an emphasis on words presumed to be unknown by readers. A further assumption is that readers need to know those words. Hence, words taught in vocabulary studies (lessons?) consist primarily of those that are on some list of important words or preselected on the assumption that readers do not know them.

Two methodological issues also contribute to the vocabulary/comprehension enigma. First is the method of instruction. The prevailing approach to vocabulary is the so-called definition approach. Teachers identify words that readers presumably need to know and then tell students to learn the definitions of those words, usually in isolation. Johnson and Pearson (1978) and Estes and Vaughan (in press) suggest that greater attention should be given to readers' prior knowledge and experiences during vocabulary instruction. Hence, readers should come to learn the concepts for word labels and concepts are learned in relation to previously known concepts. The method of instruction is, therefore, a contingent variable that can be seen as a potentially confounding variable. Ahlfors (1980) suggests that the efficacy of methodology varies depending upon the reader/word (meaning) relationships and the purpose of the instruction. A second methodological issue is the assessment procedure. Forced choice (e.g., multiple-choice, true-false) evaluations that require learners to pick out definitions reveal one trend while free recall and probe questions reveal other trends

(Anderson and Freebody, 1979). Generalizations are often drawn from specific protocols and implications rendered that may not be verifiable with other evaluation procedures.

Within this vocabulary/comprehension maze is one instructional thorn that content area teachers, in particular, face daily. Students in these classes are expected to learn new information from every reading assignment and within these assignments readers often encounter new concepts described and explained in terms that are also new and unfamiliar. The purpose of this study was to determine whether differential effects could be found between an experience based approach to vocabulary instruction and a definition based approach.

Of particular interest was the premise that the purpose of preteaching vocabulary (prior to students reading a specific selection) is to facilitate comprehension rather than to increase readers' vocabulary. Concomitant to this premise is the notion that students can best remember vocabulary terms *after* they have developed an elaborated schema for the concept that the words themselves represent. Thus, in this study we examined two approaches to preteaching key concept vocabulary and the comparative effects of each on readers' comprehension.

METHOD

Three seventh grade science classes in a suburban school district participated in this study; each class was taught by a different teacher. To offset the potential of differences among the three classes and of teacher-effectiveness, each group was exposed to three conditions (experience based instruction, definition based instruction, and no instruction). This was accomplished by using three different selections from the same textbook and the study was therefore a compilation of three instructional occasions at one-month intervals. To clarify this, the timetable is depicted as follows:

	Experience Based	Definition Based	No Instruction
Passage 1 (March)	Class A	Class B	Class C
Passage 2 (April)	Class C	Class A	Class B
Passage 3 (May)	Class B	Class C	Class A

The instructional procedures are best described in two stages, preparation and implementation. This distinction is important because the preparation for teaching is not a variable in this study; only the implementation varied. Included in the preparation stage are (1) identification of key concepts in the selection and (2) identification of vocabulary terms deemed to be essential to an understanding of those key concepts. To accomplish this, the researchers carefully read each selection and collaborated upon the delineation of the key concepts. Key concepts were defined as those concepts that teachers would prefer that students remember several years later. For purposes of contextual clarity, the key concepts were written in complete sentence form and, in sum, they represented the macrostructure for the selection. Once the key concepts were specified, the text was examined to identify those words which, if known, were most likely to facilitate readers' understanding of the key concepts or which, if not known, would impede comprehension. These key concept words were chosen only for this relative importance to understanding the key concepts. So-called difficult or potentially unfamiliar words were not selected as words to be pretaught unless they were deemed to be essential to comprehension of the key concepts. The preparation stage, therefore, determined what vocabulary would be pretaught.

Before describing the two instructional procedures, we want to insert here a brief discussion of the text from which the selections used in this study were taken. First, we should note that the selections used were chosen because they were those to be assigned at appropriate times in the classes. The teachers accommodated the researchers by coordinating their instruction so that on three specific days each class would be ready to read the same selection. These were not selections extracted and presented out of sequence or regular instructional context. When the students read the selections, the selections were those appropriate for them in their sequence of study. We should also note that the selections were of a normal length for assignments (average length 985 words) and the selections were designated by the teachers. A final observation relative to the text is of special importance. While much is being written about the difficulty of textbooks (e.g., Anderson and Armbruster, 1981) and inherent flaws within textbooks that contribute to readers' failure to comprehend, the textbook used in this study was a remarkable exception. K. Cooper's *Man and Change* (Silver Burdett, 1972) is quite well written especially in terms of coherence and unity.

When students were exposed to the definition approach, their teacher introduced the new vocabulary words via a ditto or chalkboard. The students then examined definitions for the words and any pertinent discussion, including questions, followed. The students then read the assigned passage; they could refer to the definitions while reading. After reading, each student responded to ten cued-response short answer comprehension questions.

When students were exposed to the experience based approach, the lesson began with a discussion of terms similar to the key concept words. For example, if *assassination* were a key concept word, discussion would begin with terms such as *kill* or *murder*. The teacher would gradually direct the discussion toward murder of a political figure and would conclude the discussion by introducing the term *assassination*. Each key concept word was introduced in this manner. After the lesson, students read the passage and subsequently took the test.

Students in the control group received no instruction. They read the assigned passage and took the test. The test for each selection was comprised of short answer questions to assess comprehension of superordinate, subordinate, and specific concepts (Vaughan, 1981). All tests were scored by three raters independently followed by a group meeting where differences in scoring were resolved. Each item was scored on a 0–10 basis with partial credit given as appropriate.

RESULTS

Scores for students in each condition were compiled for all three selections and analyzed via a one-way analysis of variance. A significant difference was found among the three groups for comprehension, $F(2,178) = 11.97$, $p < .001$. Tukey post hoc analysis revealed that when students were exposed to experience based instruction, their comprehension was significantly greater than when they received definition based instruction or no instruction. Vocabulary scores were similarly greater for the experienced based condition, $F(2,178) = 5.97$, $p < .01$.

The comprehension scores were also analyzed across the three conceptual levels. No significant differences existed among the conditions at the superordinate level, F (2,178) = .38. At the subordinate conceptual level, the experience based group significantly exceeded the other two conditions, F (2,178) = 5.62, $p<.01$. At the specific conceptual level, the experience based group and the control group scored significantly higher than the definition based condition, F (2,178) = 5.66, $p<.01$.

DISCUSSION

The results of this study clearly support the premise that the preteaching of vocabulary can facilitate readers' comprehension. To do so effectively, key concept words should be pretaught by initially discussing them in a context and in terms familiar to students prior to introduction of the term itself.

It may be argued that this is not a vocabulary study, per se, but instead only confirms the value of preteaching concepts prior to reading. Perhaps so, but imbedded into the experience based lesson is the preteaching of unfamiliar vocabulary in the natural context of the meanings of those terms. The terms themselves have little value for students unless the students understand the meanings of those terms. Definitions of new, technical terms often found in academic textbooks frequently convey little meaning except to those who already understand them. Hence, a definition based approach is often just so much more non-sense added to non-understanding. Perhaps a definition approach can facilitate comprehension when the definitions themselves are meaningful to students. That has yet to be verified.

In conclusion, this study supports the preteaching of vocabulary in a context meaningful for students. The passages used in this study were from a textbook that exemplifies the coherence and unity often not found in academic texts. For such clearly conclusive results to have been forthcoming with this text suggests that when students read other, less well designed texts they may benefit even more from experience based preteaching of vocabulary. Results of such a study are presently being analyzed and may lend further support to the findings reported here.

REFERENCES

Ahlfors, G. *Facilitating reading comprehension through specific vocabulary techniques.* Paper presented at annual conference of International Reading Association, St. Louis, 1980.

Anderson, R. C., and Freebody, P. *Vocabulary knowledge* (Tech. Rep. No. 136). Urbana: University of Illinois, Center for the Study of Reading, August, 1979.

Donaldson, M. *Children's minds.* N.Y.: W. W. Norton and Co., 1978.

Estes, T. H. and Vaughan, J. L. *Reading and learning in content classrooms* (2nd ed.). Boston: Allyn and Bacon, in press.

Graves, M. F. Types of reading vocabulary to teach. *Minnesota English Journal*, 9, 1978, 2–17.

Johnson, D. D. and Pearson, P. D. *Teaching reading vocabulary.* N.Y.: Holt, Rinehart, and Winston, 1978.

Tierney, R. J. and Cunningham, J. W. *Research on teaching reading comprehension* (Tech. Rep. No. 187). Urbana: University of Illinois, Center for the Study of Reading, November, 1980.

Vaughan, J. L. An initial investigation into a skills by concepts analysis of reading comprehension. In P. L. Anders (Ed.), *Research on reading in secondary schools: seventh monograph*. Tucson: University of Arizona, May, 1981, 18–25.

Vygotsky, L. S. *Mind in society*. Cambridge: Harvard Univ. Press, 1978.

MICHAEL F. GRAVES
University of Minnesota

GERALD J. BRUNETTI
St. Mary's College

WAYNE H. SLATER
University of Minnesota

THE READING VOCABULARIES OF PRIMARY GRADE CHILDREN OF VARYING GEOGRAPHIC AND SOCIAL BACKGROUNDS

From about 1910 to 1950, studies estimating the size of students' vocabularies represented an important thrust of educational research (Clifford, 1978; Lorge & Chall, 1963). Studies by Kirkpatrick (1907), Terman (1918), Seashore (1933), Smith (1941), and a number of other investigators were widely disseminated and frequently cited. In contrast, from about 1950 until the early 1970s, very little research on either vocabulary size or other aspects of vocabulary was conducted (see Dale, Razik, & Petty, 1973). In fact, as Calfee and Drum noted in their 1978 review of the reading literature, by the mid-1970s the topic of vocabulary had become a "vanishing species" (p. 217). However, over the past several years, vocabulary has again become a topic of interest to a number of educational researchers (Anderson & Freebody, 1981; Graves & Slater, 1980). Recent research has dealt with a variety of topics including the size of students' vocabularies (Dupuy, 1974; Anderson & Freebody, 1981), qualitative aspects of students' vocabularies (Graves, 1980; Mason, Kniseley, & Kendall, 1979), and the teaching of vocabulary (Barrett & Graves, 1981; Beck, Perfetti, & McKeown, in press).

The study reported here deals with the size of students' reading vocabularies. More specifically, the study examines the ability of primary grade children from three different geographic and social backgrounds to identify and assign meaning to the most frequent 5,044 words listed in the *American Heritage Word Frequency Book* (Carroll, Davies, & Richman, 1971).

The study was suggested by several conclusions Becker (1977) made after examining the successes and failures of the University of Oregon's Follow Through Program. Briefly, Becker's logical analysis of the results of the program led him to four conclusions. These were (a) that schools fail to systematically teach vocabulary, (b) that lack of vocabulary is one of the most potent factors affecting the school failure of disadvantaged children, (c) that vocabulary learning is largely additive learning, and (d) that the schools should undertake a massive, long-term program to systematically teach disadvantaged students a basic vocabulary.

Becker's conclusions are subject to a number of questions. Anderson and Freebody (1981), for example, have posed questions about the relationship between word knowledge and reading comprehension, the size of children's and adults' vocabularies, and the advisability of the sort of vocabulary program Becker proposed. Some of these questions are exceedingly complex, and the present study certainly does not provide definitive answers to any of them. However, it does provide two sorts of evidence relevant to determining the feasibility and appropriateness of systematically teaching a basic vocabulary to disadvantaged students. First, it provides a measurement of the size of students' reading vocabularies, a measurement which gives at least some indication of how many words might need to be taught. Second, it provides a direct comparison between disadvantaged students' reading vocabularies and middle class students' reading vocabularies, a comparison which directly comments on Becker's contention that disadvantaged students have severely retarded vocabularies.

METHOD

Subjects

Subjects were first, second, and third grade students from three schools. School 1 was located in a residential suburb of Minneapolis. School officials classified 2% of the students as upper class, 93% as middle class, and 5% as lower class. They further noted that 97% of the students were White and that 7% of them qualified for a free or reduced price lunch.

School 2 was located in a Colorado mountain resort and mining community of approximately 1,000 people. School officials classified 5% of the students as upper class, 90% as middle class, and 5% as lower class. They also noted that all of the children were White and that 20% of them qualified for a free or reduced price lunch.

School 3 was located in a residential, commercial, and industrial area of a major northern California city. School officials classified 20% of the students as middle class and 80% as lower class. They indicated that 85% of the students were Black, 10% were Asian, 4% were Hispanic, and 1% were White; and they noted that 99% of them qualified for a free or reduced price lunch.

Sixty-nine students from School 1, 45 students from School 2, and 79 students from School 3 participated in the study.

Materials

Materials for the study consisted of a 42 item vocabulary test and directions for giving the test. The words tested comprised a stratified random sample taken from the 5,044 most frequent words in the *American Heritage Word Frequency Book* (Carroll, et al., 1971). The strata were seven frequency blocks constructed on the basis of the words' Standard Frequency Index (see Carroll, et al.), a logarithmic index of frequency. Each successive block contained less frequent words and a larger number of words. Blocks 1–7 contained 6, 27, 77, 182, 502, 1435, and 2815 words, respectively. Taken together, these words account for approximately 89% of the running words found in reading materials used in grades three through nine. Six words were randomly selected from each block. Each item consisted of a sentence with the target word deleted and four single-word options, one of which was the target word. The test item for the word *tend* is shown below.

1. Eddie was asked to ________ the store.
A. curve   B. tend   C. shiny   D. director

Each item tested the target words' most common meaning. In addition to the test items, the test included two sample items. The directions for giving the test consisted of a foreword which described the purpose of the test and general procedures for administering it, and a script of the testing procedure itself.

Procedures and Scoring

The test was administered to the students in their regular classrooms by their teachers in mid May. When administering the test, the teacher first told students that they were going to be tested and that they should do their best but that they should not be concerned if they did not know some of the words. Next, he or she explained the testing procedure and went through the two sample items with the students. Finally, the teacher guided the students through the 42 test items. For each item, the teacher read the sentence with the target word deleted, paused about 15 seconds to let the students read the alternatives and choose an answer, and then reread the stimulus sentence, again with the target word deleted. Teachers reported that administering the test took 30–40 minutes and that the testing went smoothly.

The standard correction for guessing (Mehrens & Lehmann, 1978) was applied to students' scores prior to analysis of the data. The effect of applying the correction was to produce a more conservative estimate of students' vocabulary size than would have been obtained without it.

RESULTS

The study employed a 3 school (suburban, rural, inner city) × 3 grade (first, second, third) × 7 frequency block (blocks 1–7) factorial design with repeated measures on the last factor. The dependent measure was the 42 item test. Results were analyzed using the analysis of variance and Newman-Keuls procedures.

The analysis of variance indicated significant differences due to school, F $(2,174) = 17.49$, $p < .001$, grade, F $(2,174) = 22.79$, $p < .001$, and frequency block, F $(6,1044) = 71.02$, $p < .001$. The mean percentages of correct responses for each level of each of these factors is shown in Table 1.

The Newman-Keuls test indicated that School 3 differed significantly ($p < .01$) from both School 1 and School 2, which did not differ significantly from each other, and that first grade differed significantly ($p < .01$) from both second and third grade, which did not differ significantly from each other. The Newman-Keuls test further indicated that 16 of the 21 contrasts between frequency blocks were significant. In general, the pattern of differences was predictable. For example, all four contrasts between frequency blocks 1 and 2, on the one hand, and blocks 6 and 7, on the other were significant ($p < .01$). However, neither the contrast between frequency blocks 1 and 2 nor that between blocks 6 and 7 was significant ($p > .05$).

TABLE 1
Percentages of Correct Responses for Each School, Grade, and Frequency Block

School	School 1 = 85%	School 2 = 88%	School 3 = 68%
Grade	Grade 1 = 66%	Grade 2 = 84%	Grade 3 = 92%
Frequency Block	Block 1 = 90% Block 4 = 91% Block 7 = 73%	Block 2 = 93% Block 5 = 84%	Block 3 = 88% Block 6 = 73%

DISCUSSION

The most direct answer to the two major questions posed in the study—that of the size of primary grade students' reading vocabularies and that of the relative size of middle class and disadvantaged students' reading vocabularies—can be found by estimating the number of words out of the 5,044 words sampled that students can read. These estimates can be obtained by multiplying the number of words in each frequency block by the percentage of words correct in each frequency block. The remainder of our remarks are based on these estimates. Also, in the remainder of the discussion the results for School 1 and School 2 are combined and are interpreted as providing estimates of the size of middle class students' vocabularies, and those for School 3 are interpreted as providing estimates of the size of disadvantaged students' vocabularies.

These estimates are shown in Table 2. The answers to both questions noted above are readily apparent from the figures in Table 2. Primary grade students, both middle class students and disadvantaged students, learn to read a striking number of words. By the end of first grade, the average student probably reads in excess of 2,000 words; and by the end of third grade, the average student probably reads in excess of 4,000 words. However, although both middle class and disadvantaged students learn to read a large number of words, middle class students' reading vocabularies are substantially larger than those of disadvantaged students. For example, at the end of first grade, middle class students read about 50% more words than do disadvantaged students. Viewed from another perspective, the figures indicate that it is not until the end of the second grade that the average disadvantaged student's reading vocabulary becomes as large as that of the average middle class student at the end of the first grade. Thus, it appears that systematically teaching disadvantaged students a basic reading vocabulary would be an appropriate task. However, it would also be a very large task, the feasibility of which must await further study.

There are two cautions that should be exercised in interpreting the results. The first is that there is a definite ceiling effect in that the largest estimate possible is 5,044 words. Certainly, many students know a number of words beyond the first 5,044 most frequent ones. Also, it should be noted that the effects of this ceiling are most pronounced with the older and more able students. First grade disadvantaged students, who could read an estimated 1,791 words, probably cannot read many words beyond the first 5,044. However, third grade middle class students, who could read an estimated 4,656 words, undoubtedly know a number of words beyond the first 5,044. Additionally, it should be noted that the relatively small difference between the estimates for second grad-

TABLE 2
Estimates of Number of Words Out of the 5,044 Word Sample That Students Can Read

Grade	Type of Student	
	Middle Class	Disadvantaged
1	2711	1791
2	4038	2800
3	4656	3482

ers and those for third graders and the failure of the contrast between second and third grade scores to reach statistical significance is very likely to be due to this ceiling effect.

The other caution concerns the nature of the task presented in the test. Students were given about 15 seconds to read the alternatives for each item. Thus, students had time to decode the words, and the task did not demand automaticity (LaBerge & Samuels, 1974) or rapid lexical access (Beck et al., in press). If, as many theorists believe (see, for example, Carpenter & Just, 1981), fluent reading requires that readers automatically recognize and rapidly access the meaning of the vast majority of words they encounter, and if students did in fact consciously decode a fairly large number of the items they correctly identified on the test, our estimates do not reflect the number of words that students can read efficiently and are in that sense misleadingly high.

Factors that prompt these cautions require further explanation, and our future research in this area will attempt to avoid the ceiling effect and check for automaticity and speed of lexical access.

ACKNOWLEDGEMENT

We wish to thank Isabel Beck for her very helpful comments on the manuscript.

REFERENCES

Anderson, R. C., & Freebody, P. Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. Newark, Delaware: International Reading Association, 1981.

Barrett, M. T., & Graves, M. F. A vocabulary program for junior high school remedial readers. *Journal of Reading*, 1981, *25*, 146–150.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. The effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, in press.

Becker, W. C. Teaching reading and language to the disadvantaged—What we have learned from field research. *Harvard Educational Review*, 1977, *47*, 518–543.

Calfee, R. C., & Drum, P. A. Learning to read: Theory, research, and practice. *Curriculum Inquiry*, 1978, *8*, 183–249.

Carpenter, P. A., & Just, M. A. Cognitive processes in reading: Models based on readers' eye fixations. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1981.

Carroll, J. B., Davies, P., & Richman, B. *The American Heritage Word Frequency Book*. Boston: Houghton Mifflin Company, 1971.

Clifford, G. Words for schools: The applications in education of the vocabulary researches of Edward L. Thorndike. In P. Suppes (Ed.), *Impact of research on education: Some case studies*. Washington, D.C.: National Institute of Education, 1978.

Dale, E., Razik, T., & Petty, W. *Bibliography of vocabulary studies*. Columbus, Ohio: The Ohio State University, 1973.

Dupuy, H. J. *The rationale, development, and standardization of a basic word vocabulary test* (Vital and Health Statistics—Series 2—No. 60, DHEW Publication No. (HRA) 74-1334, National Center for Health Statistics). Rockville, Maryland: Health Resources Administration, April 1974.

Graves, M. F. *A quantitative and qualitative study of students' vocabularies*. Paper presented at the American Educational Research Association Convention, Boston, April 1980.

Graves, M. F., & Slater, W. H. *Some measurements of the depth of children's word knowledge*. Paper presented at the National Reading Conference Convention, San Diego, December 1980.

Kirkpatrick, E. A. A vocabulary test. *Popular Science Monthly*, 1907, *70*, 157–164.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293–323.

Lorge, I., & Chall, J. Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *The Journal of Experimental Education*, 1963, *32*, 147–157.

Mason, J. M., Kniseley, E., & Kendall, J. Effects of polysemous words on sentence comprehension. *Reading Research Quarterly*, 1979, *15*, 49–65.

Mehrens, W. A., & Lehmann, I. J. *Measurement and evaluation in education and psychology*. New York: Holt, Rinehart, and Winston, 1978.

Seashore, R. H. The measurement and analysis of extent of vocabulary. *Psychological Bulletin*, 1933, *30*, 709–710.

Smith, M. K. Measurement of the size of general English vocabulary through the elementary grades and high school. *General Psychological Monographs*, 1941, *24*, 311–345.

Terman, L. M. The vocabulary test as a measure of intelligence. *Journal of Educational Psychology*, 1918, *9*, 452–466.

JOANNE Y. MADISON
BONNIE A. CARROLL
PRISCILLA A. DRUM
University of California, Santa Barbara

THE EFFECTS OF DIRECTIONALITY AND PROXIMITY OF CONTEXT CLUES ON THE COMPREHENSION OF UNKNOWN WORDS

This study was designed to examine the effects of context clues on determining word meaning by their specific placements in passages relative to unknown target words. Research in the field of reading has long emphasized the importance of context as a determinant of word meaning while stressing the need for direct instruction. Beginning in 1917 with a study by Thorndike, reading has been considered a high level thinking process wherein words are stimuli that reconstruct meaning in the mind of the reader.

Some studies have attempted to classify context clues for instructional purposes. Artley (1943) and McCullough (1943) devised classification schemes, although not based on empirical evidence, while Seibert (1945) and Deighton (1959) used research data, but without specifically gleaning category evidence. Ames (1965) did not predetermine his categories but built his scheme based on introspective data obtained directly from his subjects. Quealy (1969), in replicating Ames' study, concluded that these classifications were indeed applicable.

Other researchers have studied the use of context clues by specific grade level. Biemiller (1970) and Weber (1970) both studied the effect of contextual aids used by first graders during oral reading, determining that context is used by beginning readers. Klein, Klein and Bertino (1974) found in studying fourth and sixth graders that despite performance differences between the two grades, both made effective use of context clues. Several studies were summarized by Dulin (1969) and these concluded that context clues were an effective aid in determining word meaning; that individual differences do exist in the ability to use context clues; that older students make more use of context than younger students; and that the closer the clue and the greater the amount of clues, the more effective it is.

The authors' expectations were that the context clues would prove to be an effective means of determining word meaning, when compared to no context clues; and that the location of the context clues would make a difference in their effectiveness. It was felt that the "After" condition would be more effective than the "Before" condition, since there would be more reason to utilize the clue after the unknown word had been encountered rather than prior to its appearance. The "Near" condition was expected to be more effective than the "Far" condition, because it was felt that proximity to the unknown word would increase the likelihood of the clue being recognized. Subjects in the higher grades were expected to perform more successfully than those in the lower grades, since they would have presumably read more and gained more experience in identifying context clues. Subjects high in reading ability were expected to be more successful than low ability subjects. It was theorized that there would be a developmental trend where the low ability 8th graders' performance would resemble that of the high ability 5th graders.

116

METHOD

Subjects

Thirty-four 5th graders (18 High ability and 16 Low ability), 44 8th graders (24 High and 20 Low ability) and 40 11th graders (20 High and 20 Low ability) served as subjects. At each grade level, High and Low ability readers were identified by teacher evaluation. There were five clue conditions being tested at these three grade levels: (1) No clue presented, (2) Clue presented before and distant from the target word, (3) Clue presented before and near to the target word, (4) Clue presented after and near to the target word, and (5) Clue presented after and distant from the target word.

Procedure

A list of 40 words per grade level was constructed for preliminary testing purposes. The criteria used in choosing these was their low frequency. This pretesting procedure allowed the authors to select ten unknown target words at each grade level.

In the follow-up sessions, each subject was posttested on the ten target words. Each target word was embedded without delineation in a paragraph consisting of three sentences. The target word was always situated in the second sentence. The clue conditions were varied to present the clues in positions designated as "Distant" or "Near" and "Before" or "After" the target word. "Distant" was defined as being six or more words away from the target word. The target words were all nouns, but were not restricted in terms of letters. One of the conditions presented no clues and was used as a control.

At each grade level, each of the five conditions was tested by two different words. The order of presentation of the clue conditions was randomized. After reading, with the passage in front of them, subjects were asked to make a sentence using the target word, give a definition of the target word, and identify the clue (if any) that helped them give that definition.

The scoring procedure was as follows: for sentences or definitions the subject was scored 0—no answer, 1—wrong answer, 2—somewhat correct answer, and 3—correct answer; for identifying correct clues, the subject was scored 0—wrong or no answer and 1—correct answer. The highest possible score for the sentence and definition measures of each clue was 6; the identification of the correct clue was scored as either 0 or 1.

RESULTS AND DISCUSSION

Since there was no effect for the third dependent measure (clue identification) the only dependent measures discussed will be the first sentences and the second (definitions) dependent measures. The correlation between these two measures was relatively high at .73. The results for the Sentence measure will be discussed first, then those for Definitions will be examined.

Sentences

On the first dependent measure—sentences—there was no effect for grade. There was a main effect for ability. The high ability subjects had 75% of the sentences correct while the low ability had 60% of the sentences correct. The interaction for Grade by Ability was significant ($F(2,106)=4.10$, $p<.02$), with

TABLE 1
Clue Condition by Grade and Ability for Sentence Measure

Grade	**5th**				**8th**				**11th**			
Ability	**High**		**Low**		**High**		**Low**		**High**		**Low**	
	*M**	*sd*	*M**	*sd*	*M**	*sd*	*M**	*sd*	*M**	*sd*	*M**	*sd*
No Clue	4.7	1.2	2.7	1.9	3.3	1.5	2.8	1.7	3.8	1.6	3.4	1.7
Before/Far	5.4	1.1	3.5	1.5	4.5	1.5	3.7	1.5	3.7	1.6	4.1	1.4
Before/Near	4.8	1.7	4.2	1.2	4.2	1.2	3.2	1.5	5.1	1.2	4.0	1.6
After/Near	5.3	0.8	3.3	1.8	5.2	1.2	3.1	2.3	4.2	1.6	4.1	1.2
After/Far	4.5	1.8	3.2	1.8	4.5	1.3	2.2	2.1	4.3	1.5	3.9	1.2

*out of a possible 6

the more able students maintaining their ability across grades, but with the low ability subjects fluctuating. There was a drop in performance at 8th grade and an improvement at 11th grade. There was a significant main effect for clues $(F(4,424)=6.7,\ p<.000)$.

In order to examine specific mean comparisons, three orthogonal questions were asked: (1) Is no clue better than a clue? (2) Is the "Near" position better than the "Far" position? (3) Is the "Before" condition better than the "After" condition? (1) Clue conditions were found to be much more effective than the no clue condition $(F(1,118)=24.93,\ p<.001)$, (2) The "Near" position was found to be better than the "Far" position $(F(1,118)=4.3,\ p<.05)$, and (3) There was no reliable difference between the "Before" and "After" conditions across grades.

The only substantial interaction $(p<.06)$ was Clue by Grade by Ability (Table 1). However, words were nested within grade levels so clues on each grade level were based on different sets of words. As a result we were unable to interpret this interaction. Instead we examined individual grade analyses. In the individual ANOVAS, clue was significant at $(p<.00)$ at all three grades, with orthogonal comparisons showing that no clue was less effective than any of the clue conditions. There was a main effect for Measure $(p<.00)$ at all grade levels, which showed that sentences were easier than definitions.

Definitions

Under the second dependent measure—definitions—there was a main effect for ability. The high ability subjects gave the correct answer on 63% of the definitions, while the low ability subjects had 46% of the definitions correct. There was a trend for grade, with 5th, 8th, and 11th grade having scores of 3.43, 3.25 and 3.08, respectively. However, there were no significant interactions. A significant main effect was found for clue $(F(4,424)=21.6,\ p<.00)$. As with the first dependent measure, the same three orthogonal questions were asked: (1) Is no clue better than a clue? (2) Is "Near" better than "Far"? (3) Is "Before" better than "After"? An examination of these mean comparisons showed (1) Clues were significantly better than the no clue condition $(F(1,118)=35.7,\ p<.00)$, (2) There were no significant differences found between the "Far" and "Near" positions or (3) between the "Before" and "After" positions.

Although in the overall view, grade did not appear to make a difference in performance, an analysis of individual grade results showed qualitative differences. Ability was significant at the 5th grade level where the mean score for high ability subjects was 4.49 and the mean score for low ability subjects was 2.0. Ability was also significant at the 8th grade level where the average high ability score was 4.14 and the average low ability score was 3.12. At the 11th grade level, ability made no difference. At each grade level any clue was better than the no clue condition. At the 5th grade level "Before" was better than "After" while there was no major difference between the "Near" and "Far" conditions. At the 8th grade level, "Before" was slightly better than "After". At the 11th grade level "Near" was better than "Far". The "Before/Far" condition was the most effective clue condition at the 5th and 8th grade levels while the "Before/Near" condition was the best at the 11th grade level.

Clues, in general, were found to be significant when compared to the no clue condition. When collapsed across grade levels, differences were found between the "Far" and "Near" conditions but none for the "Before" and "After" conditions. The performance of the high ability subjects showed similar patterns on both the sentence and definition measures. On these, the high ability subjects performed best under the "After/Near" clue condition. The highest scores of the low ability subjects under these measures were for the "Before/Far" clue condition. There was a trend for the higher grade levels to have lower scores than the lower grade levels. The 5th grade performed better than the 8th grade on the "After/Far" and the "Before/Far" conditions for the definition measures, and were better than the 8th grade under all clue conditions for the sentence measure. There was no apparent cause for the 8th grade drop in performance level. However, other factors may have influenced these results. The words used in the tasks were nested within grades. Those words used with the 8th and 11th grades had mean frequencies of 2.0 and 1.6, respectively, out of the Brown corpus of one million words when compared to a mean frequency of 7.5 for the 5th grade list. Thus the greater difficulty of the vocabulary may have affected their peformance. These words were selected on the principle of their difficulty to ensure a list of unknown words on which to test our clue conditions.

The classroom teachers identified the reading ability of the subjects, therefore we were unable to obtain a common standardized score on which their reading ability could be measured. The difference between the high and low ability subjects in one grade may not have been as large as within another grade. An examination of scores on the dependent measures indicated that sentences were easier than definitions. This may have been due to the fact that another sentence would easily have been patterned from the one in the test paragraph, while the task of giving a definition was more demanding.

CONCLUSIONS

We have seen that readers generally improve their use of context when it is deliberately embedded within the text. The reader could be asked to identify the meaning immediately after the target word has been presented and again immediately after the cue has occurred, to assess the influence of the specific cue on the reader's comprehension of the target word. Through on-line processing, we may be able to refine the present study and obtain greater specificity about when and what aspects of texts are used to unlock meanings of unknown words.

REFERENCES

Ames, W. S. The development of a classification scheme of contextual aids. *Reading Research Quarterly*, 1966, *2*, 57–82.

Artley, A. S. Teaching word meaning through context. *Elementary English*, 1943, *20*, 68–74.

Biemiller, A. The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly*, 1970, *6*, 75–96.

Deighton, L. *Vocabulary Development in the Classroom*. NY: Teacher's Col. Bureau of Publications, 1959.

Dulin, K. L. Using context clues in word recognition and comprehension. *The Reading Teacher*, 1969, *23*, 440–445.

Klein, G. A. and Klein, H. A. Word identification as a function of contextual information. *American Journal of Psychology*, 1973, *86*, 399–406.

Klein, H. A., Klein G. A., & Bertino, M. Utilization of context for word identification decisions in children. *Journal of Experimental Child Psychology*, 1974, *17*, 79–86.

McCullough, Constance. Learning to use context clues. *Elementary English Review*, 1943, *20*, 140–143.

Quealy, R. J. Senior high school students use of contextual aids in reading. *Reading Research Quarterly*, 1969, *4*, 512–533.

Seibert, L. A. A study on the practice of guessing word meaning from context. *Modern Language Journal*, 1945, *29*, 296–322.

Thorndike, E. L. Reading as reasoning: a study of mistakes in paragraph reading. *Journal of Educational Psychology*, 1917, *8*, 323–332.

Weber, R. M. A linguistic analysis of first grade reading errors. *Reading Research Quarterly*, 1970, *5*, 427–451.

NANCY ROSER
CONNIE JUEL
University of Texas at Austin

EFFECTS OF VOCABULARY INSTRUCTION ON READING COMPREHENSION

The relationship between vocabulary and comprehension—logical correlates in need of supporting data—underlies the three purposes of this study: 1) to determine the extent to which children *need* to learn the designated "new" vocabulary words of their basal readers; 2) to determine whether principles of teaching vocabulary—extracted from the literature and taught to teachers—would result in gains in their students' ability to identify and supply meanings for words; and 3) to determine whether vocabulary instruction has an effect on comprehension.

Relationship of Vocabulary and Reading Comprehension

Attention within the past decade in the area of reading research has been directed largely toward reading comprehension. Researchers have investigated how the reader's experiences and expectations for the printed message interact with the structure and content of the text to affect comprehension. Current reading and psychology journals, studies cited in the Annual Summary of Reading Research (Weintraub et al., 1978–79), and reports from the Center for the Study of Reading reflect the quantity of research in this area. However, one aspect of reading comprehension which seemingly contributes to how background experiences and expectations are stored, retrieved or generated has received less attention: the study of vocabulary (Manzo & Sherk, 1971–72).

There is both widespread belief and some supporting evidence for the impact of vocabulary knowledge on comprehension. As early as 1924, Hilliard found that of the factors statistically contributing to reading comprehension, intelligence ranked first and extent of vocabulary ranked second. Davis' factor analytic studies (1944, 1968) showed vocabulary knowledge to be the most highly predictive factor influencing reading comprehension, and Becker (1977), describing Project Follow Through, indicated that a lack of vocabulary is among the most powerful factors underlying economically disadvantaged student's school performance.

Marks et al. (1974) manipulated student text to determine that students who read passages containing high-frequency words scored significantly better on comprehension questions than those who read passages with low-frequency words. Marks et al. attributed this finding to the students' knowledge of high-frequency words.

Yet, recent efforts to improve comprehension scores by teaching the meanings of unfamiliar words have been either unsuccessful or marginally successful (Tuinman & Brady, 1974; Jenkins, Pany & Schreck, 1978). Tuinman and Brady pre- and post-tested 191 intermediate grade students on their knowledge of selected sets of difficult words taken from the *California Achievement Test*. Passages from the California also served as the measure of comprehension in this study. Subjects worked with self-instructional packets that required practice with definitions, fill-in-the-blanks, sentence sense, application activities, and illustration of the target word. Although all groups significantly improved their scores from pre-test to post-test, no group noticeably profited from learning the vocabulary as measured by the comprehension test. It should be noted that the lack of direct instruction and the provision of checking "keys" may have resulted in students practicing rather than learning new words—hence the lack of transfer.

In the Jenkins et al. study (1978), treatments varied in the amount of direct instruction involved. Students were taught word meanings from *context* (little direct instruction; students inferred word meanings), by being *given* the meanings (more direct instruction; teacher stated a synonym and provided a sentence) and by *practicing* meanings (heaviest direct instruction; students were drilled to mastery). Analyses supported the superiority of increased amounts of direct instruction: Practicing word meanings had greater effect than being told word meanings and both of these procedures were superior to learning through context clues alone. While some positive effects of vocabulary instruction appeared for sentence comprehension, none of the methods produced significant gains for passage comprehension.

From a review of the literature, some principles for successfully teaching vocabulary seemed to emerge:

1. *The teaching of vocabulary must be direct and systematic.* (Gray, 1938; Dale et al., 1971; O'Rourke, 1974; Jenkins et al., 1978; Becker, 1977; Dale & Milligan, 1970.)
2. *Instruction should help the learner draw the relationships between words to be learned and the learner's own background, needs, and interests.* (Gipe, 1978–79; Addy, 1941; Clifford, 1978; Manzo & Sherk, 1971–72; Carroll, 1964; O'Rourke, 1974.)
3. *Instruction should provide in-depth work with vocabulary words rather than cursory attention.* (Deighton, 1959; Johnson & Pearson, 1978.)

While strategies can be identified, there is little evidence to suggest that they are being employed in any systematic way. Durkin (1978–79) noted the absence of attention to vocabulary instruction when, after observing over 4,000 minutes of reading instruction in 24 classrooms, she found only 2.6 percent of the instructional time devoted to word meanings—whether in teaching, review, or in application. This finding corroborated that of Austin and Morrison who concluded that, while there was considerable stress placed on the development of word recognition techniques in classrooms, there was little corresponding attention to the meanings of words. As further corroboration, we observed the reading lessons of 12 classroom teachers in grades one through five in order to determine the extent of their vocabulary instruction. We observed a total of 40 reading lessons over a 3 month span, or approximately 1200 minutes of reading instruction. The total time spent on instruction in word meaning was 65 minutes. The mean was 1.67 minutes per lesson, with a range from 0 to 12 minutes. But more important, the mode was zero. Further, we asked these same teachers to keep a daily journal of their attention to vocabulary instruction throughout the day in content area subjects. Attention to word-meaning instruction seemed minimal or missing.

Explanations for inattention to vocabulary instruction are few. Logic suggests that teachers have too little time or, alternatively, they perceive that children do not need vocabulary instruction, or need it less than they do other skills. The literature seems clearer on other instructional questions, e.g., instruction in vocabulary tends to result in students learning vocabulary, although this learning is not necessarily judged using realistic tasks. Third, the relation between vocabulary and comprehension, seemingly a close one, is not instructionally untangled.

METHOD

Subjects

Sixty-six children were involved in the study. Approximately half were from middle reading groups, and half from low reading groups. Similarly, lower grade children (grades 1 and 2) and middle grade children (grades 3, 4, and 5) were equally represented. The children were from 12 different classrooms in 7 lower to middle class schools in Austin, Texas. Their instructors for the experiment were their normal classroom teachers. These 12 teachers were enrolled in a Masters of Education program and participated in the study to fulfill a course requirement.

Before instruction began, informal reading inventories, built from the classroom's basal series, were administered to validate the students' instructional level and accuracy of placement. Reading groups selected for the study had all children reading in basals at their instructional level. Students from the middle reading groups were receiving instruction in a basal equivalent to their grade placement. While students in the low reading groups were also placed at their instructional levels, their basals were on one level below their current grade placement.

Materials

Vocabulary words tested in the study were the "new" words in a story as identified by the American Book Company or Scott-Foresman basal reader teachers' editions (between 5 and 10 words per story). Each reading group was tested on the vocabulary from 2 consecutive basal stories in the basal they were currently reading. The stories were selected because they were the ones that normally would have been introduced next in class.

Design

Because we were interested in results generalizable to actual classroom situations, we chose to test the effects of vocabulary instruction on intact reading groups, with the normal classroom teacher as the instructor.

All subjects in the experiment participated in both an instruction and no instruction condition. Order of the conditions was randomized. In each condition there were identical pre- and post-tests on word identification and meaning knowledge. A 2×2×2 design was used with a within subjects factor of Instructional Condition (Instruction or No Instruction) and between subjects factors of Reading Group (Middle or Low) and Grade (Grades 1 & 2 vs. Grades 3, 4, & 5). There were 3 dependent measures: % Word Identification Gain (pre- to post-test), % Meaning Gain (pre- to post-test) and % Comprehension Questions answered correctly.

The comprehension questions were developed to tap literal grasp of the passage in which the new vocabulary word first appeared. For example, if the new word was *measure*, a question and probes were framed to gauge whether a child had understanding of the passage in which the word measure appeared.

T: What did Billy's mother do on Billy's birthday?
C: She measured him on the wall.
T: What did Billy's mother want to know?
C: How much Billy had grown.

Through this procedure, judgments of comprehension were subjective and accuracy determinations were left to the teacher's discretion. In each instance—for each child—the determination of comprehension was done individually and the decision made was whether or not the question was answered accurately. There was no partial scoring—an answer was either judged correct or incorrect.

Procedure

All 66 children were pretested individually both in ability to identify words and to supply meanings for the new vocabulary words in the next consecutive basal reader story to be encountered. All words were identified by the teacher's editions as being the "new words." Pre-test procedures involved asking individuals to first pronounce vocabulary words displayed on 3" by 5" index cards. Next, the word was pronounced (without the card present), and the child was asked what the word meant. Testers were given training in how to probe the child's responses in order to make accurate judgments of the child's grasp of a word's meaning.

All teachers were taught a method of teaching vocabulary based on the principles described in the literature review (direct teaching, drawing relationships between vocabulary and the readers, and in-depth treatment). The following procedure served as the instructional steps for presenting new words:

1. Teachers provided a chart containing sentences that gave clues to the new vocabulary words. Each sentence contained a blank where the target word should be. Next to the chart (along the chalkboard ledge) the vocabulary word cards were displayed.
2. Teachers read the sentences aloud.
3. Children were encouraged to decide which new word fit the context. Correct suggestions as well as incorrect ones were tried in the sentences, and children were asked to make judgments about sentence sense. Children were also encouraged to tell what else they knew about a word. Teachers expanded on definitions by attempting to relate the words to the children's lives (e.g., "Remember when we made pudding in class? We had to put in two cups of milk. Mark *measured* the milk. He poured it carefully into the cup.").
4. A child who correctly identified the new word was asked to tell some ways he accomplished this. This step was designed to ensure that children extracted strategies from their knowledge of language and reading for use in determining word meanings independently.
5. The sentences were reread with the correct word physically inserted.
6. After all the words had been inserted into spaces, the words were reviewed in random order.
7. The procedure varied in time demands from 8 to 12 minutes. The emphasis of the vocabulary instruction was on meaning, though a by-product of seeing the words might be increased word identification skill.

The groups were randomly assigned to an instruction or no instruction condition for the first story. For the next story lesson, the groups were reversed. In both conditions (instruction or no instruction), post-testing included an opportunity to 1) identify the new word; 2) supply a meaning or offer an example of that word; and 3) answer orally presently literal or textually explicit questions about the story, the answers to which involved the words that were taught.

Analysis of Data

Data were originally scored as percentages (i.e., percent words correct out of total words tested). A weighted means ANOVA was used to analyze the data on the 3 dependent measures—Word Identification, Meaning, and Comprehension. Word Identification and Meaning Identification are treated as gain scores (the percent difference between the pre- and post-tests in the Instruction and No Instruction conditions). However, all the data were logarithmically transformed from percentages before they were entered into the ANOVAs, using a logit transformation.

RESULTS AND DISCUSSION

F-values, and means and standard deviations for main effects are reported in Table 1. Discussion is based on the questions which guided the study.

What proportion of "new" words in a basal story can students identify and supply meanings for prior to reading the story? Results of the study indicated that, prior to reading their basal stories, both low and middle group children recognized only about half the words specified as "new" words by the teachers' editions. Yet, intermediate grade students (Grades 3, 4, 5) identified 72% of the "new" words while beginning readers (Grades 1 & 2) identified only 42.8% of the "new" words prior to reading the basal story.

Middle reading group children could supply the meanings for 64% of the "new" words prior to story reading while low reading group children supplied 48% of the "new" word meanings. Grade levels appeared about the same in knowing word meanings prior to story reading (Grades 1 & 2—55.8%; Grades 3, 4, 5—58.9%).

Do children show gains from vocabulary instruction in word recognition and word meaning knowledge? Vocabulary instruction did not produce significant gains on post-test performances in word recognition. This is not surprising in that the focus of the instruction was on meaning. Interestingly, students showed gains on the post-tests after simply reading the "new" words in the context of the story. Beginning readers (Grades 1 & 2), regardless of group membership, showed significantly more gains overall than did intermediate grade readers (Grades 3, 4 & 5).

Specific vocabulary instruction prior to reading a story, however, significantly affected the post-test performances of meaning knowledge. While children appear to be successful learning to identify words by simply reading those words in the basal story context, they do not appear to be as successful learning the meaning of words through story context. Rather, direct instruction in word meaning prior to story reading appears especially helpful. In particular, there is an indication that low reading group children profit from direct word meaning instruction (Reading Group × Instruction Condition $F(1,63)=3.6$, $p<.06$, Table 1). At intermediate levels, similar mean gains were achieved regardless of

TABLE 1
Analysis of Variance for Three Dependent Measures

	F	Means	SD
Word Identification[a]			
Reading Group	.7		
Middle		34.9	20.4
Low		30.5	26.9
Grade	6.6*		
1, 2		40.7	23.7
3, 4, 5		18.5	14.5
Instruction	3.8		
Instruction		35.0	24.0
No Instruction		31.0	23.1
Grade × Instruction	.8		
Rdg. Grp. × Instruction	.9		
Rdg. Grp. × Grade	4.4*		
Rdg. Grp. × Grade × Instruction	.1		
Meaning[a]			
Reading Group	.0		
Middle		28.1	23.9
Low		30.3	21.4
Grade	.6		
1, 2		29.4	25.6
3, 4, 5		28.4	16.9
Instruction	10.0**		
Instruction		32.6	23.6
No Instruction		25.4	22.3
Rdg. Grp. × Instruction	3.6		
Grade × Instruction	.4		
Rdg. Grp. × Grade	4.3*		
Rdg. Grp. × Grade × Instruction	.1		
Comprehension[b]			
Reading Group	4.2*		
Middle		81.6	19.1
Low		76.7	21.5
Grade	4.0*		
1, 2		85.8	18.0
3, 4, 5		74.7	24.2
Instruction	2.8		
Instruction		84.5	16.9
No Instruction		76.1	25.2
Rdg. Grp. × Instruction	4.7*		
Grade × Instruction	.1		
Rdg. Grp. × Grade	8.1**		
Rdg. Grp. × Grade × Instruction	8.2**		

* $F(1,63)=3.99$, $p<.05$

** $F(1,63)=7.05$, $p<.01$

[a] Means are percent gains between pre- and post-tests.

[b] Means are mean percent correct.

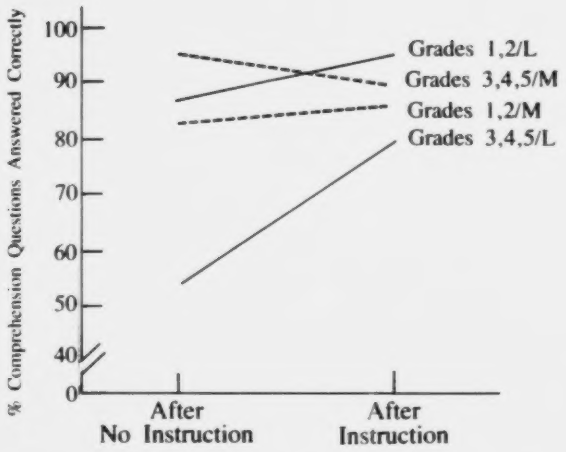


FIGURE 1: Interaction of Reading Group × Grade Level × Instruction.

group membership, but younger readers in the low groups made significantly greater gains than both their peers in the middle group and intermediate level low group children (Reading Group × Grade Level $F(1,64)=4.3$, $p<.05$, Table 1). Overall, direct meaning instruction seems beneficial and, for low reading groups, perhaps essential.

Is there an effect of vocabulary instruction on comprehension (as measured by answers to textually explicit questions, the answers to which involve the vocabulary that was pre-tested)? The vocabulary instruction we employed emphasized growth in knowledge of word meanings. We were interested in whether this vocabulary instruction would translate into improved story comprehension. While there was no significant main effect of vocabulary instruction, there were significant interactions which parallel our previous findings that low reading groups, especially those at intermediate levels, are the most needy of direct vocabulary instruction (Figure 1). Without vocabulary instruction, the comprehension of low ability upper grade students suffers. It should be noted that it is meaning instruction, rather than its possible by-product of increased word recognition skill, that appears to account for this result. For low ability, upper grade students simply reading the text improved word recognition as much as instruction: (Mean word recognition after no instruction = 89% and mean word recognition after instruction = 90%).

CONCLUSIONS

Thus, we have found that children do not appear to know all the words prior to instruction, yet that instruction itself has less effect on word identification than does grade placement. Lower grade children make significantly greater gains in word recognition regardless of reading group membership. This may be attributed to the relative difficulty of the words intermediate grade children were asked to pronounce or that beginners are somehow attuned to word identification because of the instruction emphasis of their program.

Growth in word meaning is expectedly affected by instruction. In addition, lower grade children assigned to low reading groups made the most significant gains in meaning knowledge. Again, this phenomenon can be related to the group's relative need for vocabulary.

Finally, instruction in vocabulary appears to have had the greatest effect on improving the comprehension of the intermediate grade children assigned to low reading groups. Intermediate grade children in middle ability groups were least dependent on vocabulary instruction for making gains in comprehension. They were able to answer textually explicit questions with a good deal of accuracy in both instruction conditions. Lower grade, low reading group children also improved as a result of instruction. Apparently, if teachers decide to spend time on vocabulary instruction, the comprehension of their low ability children will improve. As text difficulty increases for those low group children—as for those placed in intermediate readers—the effects on word meaning instruction will be most pronounced.

REFERENCES

Addy, M. L. Development of a meaning vocabulary in the intermediate grades. *Elementary English*, 1941, *18*, 22–26, 30.

Becker, W. C. Teaching reading and language to the disabled—what we have learned from field research. *Harvard Educational Review*, 1977, *47*, 518–543.

Campbell, D. J., & Stanley, J. C. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally, 1966.

Carroll, J. B. Words, meanings and concepts. *Harvard Educational Review*, 1964, *34*, 178–202.

Clifford, G. J. Words for schools: the applications in education of the vocabulary researches of Edward L. Thorndike. In P. Suppes (Ed.), *Impact of Research on Education: Some Case Studies*. Washington, D.C.: National Academy of Education, 1978, 107–198.

Cohen, J. & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1975.

Dale, E., & Milligan, J. L. Techniques in vocabulary development. *Reading Improvement*, 1970, *1*, 1–15.

Dale, E., & O'Rourke, J. *Techniques of teaching vocabulary*. Palo Alto, CA: Field Educational Publications, 1971.

Davis, F. B. Fundamental factors of comprehension in reading. *Psychometrika*, 1944, *9*, 185–197.

Davis, F. B. Research in comprehension on reading. *Reading Research Quarterly*, 1968, *3*, 499–545.

Deighton, L. C. Developing vocabulary: Another look at the problem. *English Journal*, 1960, *49*, 82–88.

Durkin, D. What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 1978–1979, *14*, 481–533.

Gipe, J. P. Investigating techniques for teaching word meanings. *Reading Research Quarterly*, 1978–1979, *14*, 624–644.

Gray, W. S., & Holmes, E. *The development of meaning vocabularies in reading*. Chicago: University of Chicago Press, 1938.

Guszak, F. J. *Diagnostic reading instruction in the elementary school*. New York: Harper & Row, 1978.

Harber, J. R., & Bryen, D. N. Black English and the task of reading. *Review of Educational Research*, 1976, *46*, 387–405.

Hernandez, N. G. Variables affecting achievement of middle school Mexican-American students. *Review of Educational Research*, 1973, *43*, 1–39.

Hilliard, G. H. Probable types of difficulties underlying low scores in comprehension tests. University of Iowa *Studies in Education*, 1924.

Jackson, J., & Dizney, H. Intensive vocabulary training. *Journal of Developmental Reading*, 1963, *6*, 221–229.

Jenkins, J. R., Pany, D., Schreck, J. *Vocabulary and reading comprehension: instructional effects*. Technical Report No. 100. University of Illinois, Champaign-Urbana, IL: Center for the Study of Reading, 1978.

Johnson, D. D., & Pearson, P. D. *Teaching reading vocabulary*. New York: Holt, Rinehart & Winston, 1978.

Manzo, A. V., & Sher, J. K. Some generalizations and strategies for guiding vocabulary learning. *Journal of Reading Behavior*, 1971-72, *4*, 78–79.

Marks, C. B., Doctorow, M. J., Wittrock, M. C. Word frequency and reading comprehension. *Journal of Educational Research*, 1974, *67*, 259–262.

O'Rourke, J. P. *Toward a Science of Vocabulary Development*. The Hague: Mouton, 1974.

Petty, W. T., Herold, C. P., Stoll, E. *The state of knowledge about the teaching of vocabulary*. Champaign, IL: National Council of Teachers of English, 1968.

Tuinman, J. J., & Brady, M. E. How does vocabulary account for variance on reading comprehension tests? A preliminary instructional analysis. In *Interaction: Research and Practice for College-Adult Reading. Twenty-third Yearbook of the National Reading Conference*, 1974.

CONNIE A. BRIDGE
University of Kentucky

BRENDA BURTON
Stonewall Elementary School
Lexington, Kentucky

TEACHING SIGHT VOCABULARY THROUGH PATTERNED LANGUAGE MATERIALS

Frank Smith (1978) contends that children learn to read by reading and that the function of teachers is not so much to teach reading as to help children read. The dilemma lies in finding materials for beginning readers to use to practice reading even as they are learning to read. Kenneth Goodman (1976, p. 5) asserts that "one of the things that's most important in early reading material, maybe the only really important thing, is its predictability." That is, children must be able to find what they expect to find in early reading materials in terms of language, ideas, and concepts. In an earlier article, the first author (Bridge 1979) explained why the most predictable materials for novice readers are the language experience stories which children dictate for themselves and the patterned language or structured language books which have repetitive structures that enable children to anticipate the next word or line or episode. These repetitive structures also provide children with many opportunities to recognize words in a dependable context. The continued exposure to high frequency words in these predictable materials helps to develop a sight vocabulary of words that can be recognized in a variety of contexts.

One of the major goals of the beginning reading materials in most basal reader series is also that of building up a sight vocabulary of high frequency words which children can recognize instantly. In an attempt to help children learn to recognize these words, the authors of preprimers and primers use these same few words over and over in short, stilted sentences. Sentence length is rigidly controlled as studies of readability have shown that, as a general rule, shorter sentences are easier to read than longer sentences. However, when carried to an illogical extreme, the strict control of sentence length and vocabulary produces unnatural language and plotless strings of disconnected sentences which may actually interfere with beginning readers' ability to predict the next word or phrase because the syntactic patterns do not match the children's own language patterns and the meaningless groups of sentences do not match their schematic expectations for well-structured stories in which events unfold in a natural, logical way.

The purpose of this study was to determine whether sight recognition of high frequency words could be accomplished as effectively or more effectively through the use of patterned language materials as through the use of the readiness materials in a major basal reader series in which the explicitly stated purpose is the instant recognition of a core of high frequency sight words. These initial reading materials, *Getting Ready to Read* published by Houghton-Mifflin, comprise part of the readiness materials in the adopted series in the district in which the study was conducted. The approach used in these materials is labeled as an oral context-rebus approach by the authors.

The specific research question to be answered was:

Will children using the patterned language materials be able to recognize a greater number of target and nontarget high frequency words than the children using the rebus approach?

METHOD

Subjects

The subjects were 16 kindergarten children in an upper middle class school in Lexington, Kentucky, which ranks above national norms on standardized reading tests. The study was conducted during late April and May. The children who participated in the study comprised the lowest groups in the morning and afternoon kindergartens taught by the second author. They were the only children who had not already completed the Houghton-Mifflin *Getting Ready to Read* materials. The morning group, consisting of five girls and two boys, received instruction using the patterned language materials. The afternoon group was made up of five girls and four boys and used the *Getting Ready to Read* rebus materials.

Materials

The authors of Houghton-Mifflin have selected sixteen of the top 500 words identified in *The American Heritage Word Frequency Book* as the words occurring with highest frequency in samples of reading matter used in public, private, and parochial schools. These 500 words accounted for 90 percent of the approximately one and a half million running words which appeared in the third and fourth grade materials in the sample. The 16 target words out of these top 500 words accounted for more than 28 percent of those running words. The authors designed the *Getting Ready to Read* materials to teach these 16 high frequency words so that when students start reading the first preprimer they will be able to recognize most of the words in that preprimer and use those words as a contextual base for decoding the other words introduced in the preprimer.

Three patterned books were selected which contained among the other words in the text the same sixteen high frequency words introduced in *Getting Ready to Read*. The books were *Brown Bear, Brown Bear, What Do You See?* by Bill Martin, Jr. (1971a); *One, Two, Three, Four* by Bill Martin, Jr. (1971b); and *A Bug in a Jug and Other Funny Rhymes* by Gloria Patrick (1974).

Procedure

Prior to the beginning of the study, the children were given a pretest on recognition of the upper and lower case letters of the alphabet. A preliminary analysis indicated that the two groups did not differ significantly in their knowledge of either upper or lower case letter names. They were also pretested on 40 high frequency sight words which included the 16 target words taught in *Getting Ready to Read* plus 24 other words taken from a list of high frequency words (Petty, Petty, & Becking, 1974). A post-test of these 40 high frequency sight words was administered at the end of the study. Both groups were instructed by their own teacher 15 minutes per day for twenty-four days over a six-week period, a total of six hours of instruction.

The oral context/rebus sentence group received the prescribed lessons in the *Getting Ready to Read* practice book. Once the children were introduced to the words through letter-sound associations and oral context, they were given a great deal of recognition practice in rebus sentences and stories. This program begins with a teacher-provided oral context and the visual presentation of the high frequency target word in isolation. Students are first encouraged to use letter-sound associations and the oral context to figure out the new word. Then they are asked to identify the word in a list of other isolated words and then read it in short rebus sentences (5–7 words long) and then in rebus stories (2–6 sentences long). Attention is initially directed to the grapho-phonic characteristics of the word with some contextual support (syntax and semantics) provided orally and then visually in the rebus sentences and stories.

The patterned language group followed a sequence of lessons which combined the suggestions of Bridge (1979), an adaptation of Cunningham's (1979) procedures for developing sight vocabulary through structured language experience, and Martin & Brogan's (1971) guidelines for using patterned books. The sequence of lessons was:

1. The teacher read the book to the group and then read it again inviting them to join in when they could predict what was coming next. The group was then divided into two groups for choral reading of the book, changing parts for the second choral reading.
2. Teacher and students read the story from the book. Then they read it again from a chart without the picture cues contained in the book.
3. Teacher and students read the story from the chart. The students were given sentence strips containing the lines of the story. The students matched the strip with the corresponding line of the story, placing the strip under that line on the chart.
4. The students read the story from the chart together. They were then given individual word cards in order which they placed under the matching words on the chart.
5. Students read the story from the chart together. The teacher placed individual words from sections of the story out of order at the bottom of the chart. As the teacher named individual words, students located the words and placed them in order. This procedure was followed for each section of each story.

This procedure proceeds from whole to part. The children hear and read the whole story first before they are asked to work with individual sentences and words. They always have the contextual support of the whole text as they match the sentence strips and word cards. Thus, they are always dealing with large chunks of meaning (semantics) and natural sounding language (syntax) as their attention is directed to the significant features of the words (graphophonics).

RESULTS

To determine whether or not the children using the patterned language materials would be able to recognize a greater number of high frequency words than the oral context/rebus group, three t-tests were run to test the differences between the mean number of the 40 total high frequency words learned during the study, the mean number of the 16 target words learned, and the mean number of nontarget words learned from the remaining 24 high frequency words. There was no significant difference in total words learned and no difference in number of target words learned; however, the patterned language group learned significantly more nontarget words. $t=2.35$, df 7,14, $p.<.05$ (See Table 1).

TABLE 1
Mean Gain Scores for Total High Frequency Words, Sixteen Target Words, and Nontarget Words

	Total		Target		Nontarget	
	M	S.D.	M	S.D.	M	S.D.
Group 1	5.000	(3.559)	3.000	(2.160)	*2.000	(1.915)
Group 2	4.111	(1.764)	3.889	(1.453)	0.222	(0.667)

Group 1—Patterned Language
Group 2—Rebus

*p.<.05

The results indicate that the students using the patterned language materials were able to recognize slightly more total high frequency words than the oral context/rebus group which in turn learned slightly more target words. Neither difference was significant. Because of their opportunity to recognize more words within meaningful contexts, the students reading the pattern language books were able to learn significantly more of the nontarget high frequency words. Perusal of the 40 high frequency words indicates that most of them are structure or function words, such as articles, pronouns, prepositions, conjunctions and auxiliary verbs. No attempt was made to assess the children's recognition of other important form class words, such as adjectives and nouns which were repeated frequently in the patterned language materials, specifically color words, animal names, cardinal and ordinal numbers and other high frequency nouns. Because of the strict control of the rebus materials, the other group of children had no opportunity to read these words in meaningful contexts during the instructional sessions. Thus, it is doubtful that they had acquired very many of these words. The differences between the two groups in acquisition of incidental words would probably have been greater if these other form-class words had been included in the sight word recognition tests.

DISCUSSION AND IMPLICATIONS

It appears that the students in this study learned the high frequency target sight words as well in the patterned language materials as they did in the more restrictive rebus sentences and that they learned more nontarget high frequency words in the more meaningful contexts of the patterned books. Before results can be generalized to students of varying ability levels and socioeconomic status, it will be necessary to replicate the experiment with other populations of beginning readers. No attempt was made to assess incidental learning of other high frequency form class words or affective factors such as the children's concept of the purposes of print and their view of the reading process. Natural language and meaningful contexts should create a more positive attitude toward learning to read as children experience the values of "wholebooksuccess" (Martin, 1971). It is important to note that the oral context/rebus sentence approach does provide some semantic and syntactic support for the acquisition of sight vocabulary. It would also be interesting to compare the patterned materials to more synthetic approaches to word recognition in which children are asked to deal with letter-sound relationships without access to the semantic and syntactic information provided by context.

REFERENCES

Bridge, C. Predictable materials for beginning readers. *Language Arts, 56,* 503–507.

Cunningham, P. Beginning reading without readiness: Structured language experience. *Reading Horizons,* 1979, *19,* 222–227.

Goodman, K. S. *Reading: A conversation with Kenneth Goodman,* Glenview, IL: Scott Foresman and Company, 1976.

Lewis, J., Harrison, M. L., Durr, W. K., & McKee, P. *Getting ready to read.* Boston: Houghton-Mifflin, 1979.

Martin, B. *Brown bear, brown bear, what do you see?* New York: Holt, Rinehart, and Winston, 1971a.

Martin, B. One, two, three, four. *In Sounds of numbers.* New York: Holt, Rinehart, and Winston, 1971b.

Martin, B. & Brogan, P. *Teacher's guide to the instant readers.* New York: Holt, Rinehart, and Winston, 1971.

Patrick, G. *A bug in a jug and other funny rhymes.* New York: Scholastic Book Services, 1974.

Petty, W. T., Petty, D. C., & Becking, M. F. *Experiences in Language: Tools and techniques for language arts methods.* Boston: Allyn and Bacon, 1974.

Smith, F. *Understanding reading: A psycholinguistic analysis of reading and learning to read* (2nd Ed.). New York: Holt, Rinehart, and Winston, 1978.

BEVERLY COX
ELIZABETH SULZBY
Northwestern University

EVIDENCE OF PLANNING IN DIALOGUE AND MONOLOGUE BY FIVE-YEAR-OLD EMERGENT READERS

This paper compares instances of conversation and storytelling taken from case studies of three kindergarten children judged to be high, moderate, and low in emergent reading ability. The case studies come from a project in early literacy called "Beginning Readers' Developing Knowledges About Written Language," or BRDKAWL. The goal of the project is to provide a theoretically-based description of the transition from pre-reading to reading. The description includes many facets of the child's knowledges about oral and written language, including conversation, storytelling, dictation, writing, re-reading, and editing.

Conversation and storytelling, as structured in this project, differ along dimensions related to the distinction between dialogue and monologue. We propose that the ability to distinguish between the oral language modes of dialogue and monologue is one indication of the child's growing knowledge about written language. We further suggest that a characteristic of thought which Vygotsky (1962) called inner speech underlies both dialogue and monologue and often intrudes into these forms, particularly for young children.

Inner speech refers not only to the covert cognitive machinations that we typically call thinking; it also includes all uses of speech to facilitate thinking, particularly planning. We often excuse such speech: "Sorry, just thinking out loud." Inner speech, according to Vygotsky, is both natural and vital to cognitive and linguistic growth. The purposes it attains are refinement, practice, and learning of control through feedback. When externalized, inner speech requires significant translation in order to communicate effectively. Inner speech's form is abbreviated and highly contextualized and its function, rather than being communication, is planning.

Conversation is a form of dialogue, or speech with others, in which responsibility for the total speech event is shared between partners. Blank (1973), Bruner (1975), Cherry (1979), and Wertsch (1979) have demonstrated that the ability to carry the burden in conversation is gradually assumed by the child through interaction with an adult, usually the mother. Wertsch's research suggests that the adult acts as a metacognitive guide for the child. The caretaker models thinking through dialogue with the child and the child gradually becomes able to think internally using the steps which the adult has modelled. The adult provides verbal cues that serve the same planning function as the child's own inner speech will serve when s/he is older.

Storytelling, as defined in the BRDKAWL paradigm, is a form of monologue. In contrast to dialogue, monologue can be considered to be speech for others. Monologue can be oral or written, but in either case the composer needs to make a communication clear either to an absent or a non-interactive audience. Sulzby (1980) suggested that both reader and writer need an internal model of the reader-author relationship and that the ability to sustain a monologue furnishes evidence of how this model is developing.

In studying kindergarteners, Sulzby (1981, 1982) found that children who show less emergent reading ability tend to modify speech situations requiring monologue toward interactive, conversational speech; it is as if the child's plan for composition has to be externally provided. These conclusions were based upon differences in children's speech in storytelling and story dictating conditions and upon comparisons with children's emerging abilities to read their dictated and handwritten stories.

Thus it is suggested that, within mainstream literacy cultures, oral language forms can intrude into tasks requiring written language forms and that such intrusions are related to the literacy knowledge of the child in relation to the given task. The child more mature in literacy can make clearer distinctions between linguistic modes and tasks and can make these distinctions more voluntarily. In contrast, the less mature child makes distinctions that are less clear and upon request has little success without continual aid. Sulzby has suggested that the aid which less mature children elicit tends to modify the linguistic situation, e.g., from the mode of storytelling toward conversation. More mature children, on the other hand, tend to modify modes to show characteristics of written language forms. Such modifications may also include intrusions of characteristics of inner speech; we would expect more instances of inner speech to intrude in monologue for mature children. For children with less linguistic maturity, inner speech instances should be evident in dialogue as well.

In this paper, two modes, conversation and storytelling, were examined for the child's effectiveness in maintaining dialogic and monologic forms. Samples of each mode were analyzed for intrusions of inner speech. Both mode effectiveness and characteristics of inner speech were compared with the literacy development of the child being studied. These questions were addressed descriptively: (1) How well can the child perform as an equal participant in conversation (dialogue)? (2) How well can the child produce autonomous text (monologue) that is an uninterrupted cohesive entity not requiring questions from its intended audience? (3) What is the nature of inner speech intrusions in the child's dialogue and monologue?

METHOD

Subjects

The subjects were three kindergarten children from a school in a suburb north of Chicago, Illinois. Selection for case study was based upon results of the first of a series of five BRDKAWL studies. The children were chosen to vary in overall emergent reading ability:[1]

Code name:	Emergent Reading Ability:	Age as of 12/80:	Sex:
Betsy	High	5–9	Female
Andrea	Moderate	5–7	Female
Mike	Low	5–8	Male

Data Source

The data come from the five BRDKAWL studies that extended across the 1980–81 school year. Complete transcripts were made from tape-recorded sessions in which children were asked to perform a recurring set of tasks: storytelling, dictation, story writing, with re-reading and editing of the two written forms. Additionally, instances of conversation and interrogation were built into the structure of all five studies so that protocols included many instances of both dialogue and monologue. From these instances, the first author selected 92 samples judged typical of each child's dialogue and monologue attempts. The two authors then selected subsets to be analyzed.

Analysis

Dialogue and monologue effectiveness. Five sets of samples consisting of one sample per child were presented to two independent judges for ranking according to the effectiveness of the child's participation in dialogue or maintenance of monologue. The three dialogue sets came from Study I (November) and Study III (March). The two monologue samples came from Study II (January–February) and Study IV (April–May).

The two judges differed in linguistic training and directions for the effectiveness judgments were devised to suit their training. To determine whether "effectiveness" could be judged by an ordinary person acquainted with children, we asked a naive judge simply to take each set and rank it according to how effective the child was in maintaining either dialogue or monologue. After ranking, she was asked to describe the bases used in making the decisions. These bases were compared with those formulated for the experienced judge; the bases proved compatible.

The second judge, trained in linguistics, was given the following criteria for the two modes of speech. Dialogue samples were ranked according to the child's effective use of turn-taking, specification of referents, and event sequence. The child would not be judged effective if the adult had to interrogate the child for needed information or had to bear most of the burden of maintaining dialogue. Effective monologue was judged by how well the child composed an autonomous unit which could be understood without external intervention such as questions and which maintained an appropriate level of decontextualization. Each set of three samples was ranked separately. Judges agreed 87%, for 13 of 15 rankings. Rankings by the experienced judge appear in Table 1.

Characteristics of inner speech. Two trained judges reviewed all samples used in the effectiveness rankings and selected instances which could be considered intrusions of inner speech. Instances fell into two categories: "Planning" and "Other."[2]

Individual and developmental description. Case studies for each of the three children were based upon the five studies described above, classroom observations, and interviews with parents and teacher. Comments using case study information are woven into the results and discussion sections. One conclusion from the case studies that is relevant is that the three children maintained their overall standing in emergent reading abilities through the year.

RESULTS

Dialogue and monologue effectiveness. Table 1 shows the effectiveness rankings for the dialogue and monologue samples compared with each child's rankings on emergent reading ability. Betsy, the child ranked highest in emergent reading ability, performed at a steady level in both modes. In contrast, Andrea (moderate) and Mike (low) exhibited differential competencies. Mike was ranked slightly more competent in dialogue than was Andrea and both were

TABLE 1
Rankings of Dialogue and Monologue Samples Compared With Emergent Reading Ability

Child	Emergent Reading Ability Rank	Dialogue Ranks				Monologue Ranks		
		Fall 1	Fall 2	Spring	Median	Winter	Spring	Median
Mike	3[a]	2	3	2	2	3	3	3
Andrea	2	3	2	3	3	1	2	1.5
Betsy	1	1	1	1	1	2	1	1.5

[a] 3 equals the ranking of least ability.

judged less competent than Betsy. However, in monologue, Mike was clearly ranked last while Andrea approached Betsy's performance.

Characteristics of inner speech. Thirty-four instances of inner speech were agreed upon by the two raters. Table 2 shows the frequency distribution. While these frequencies were not corrected for length, it can be seen that both Betsy and Mike produced many instances. However, Mike employed inner speech almost equally in monologue and dialogue, while Betsy concentrated her inner speech in monologue. A review of the protocols revealed that Mike relied more on dialogue with the examiner in order to produce the sample labelled monologue, while Betsy used inner speech in the monologue itself. In other words, Mike had the inner speech function accomplished for him by the examiner and Betsy produced her own. Thus, too, his instances of inner speech in monologue are under-represented in this tally.

TABLE 2
Frequency Data:
Inner Speech Instances in Dialogue and Monologue Samples

Child	Emergent Reading Rank	Dialogue			Monologue		Totals
		Fall 1	Fall 2	Spring	Winter	Spring	
Mike	Low	3	2	1	3	2	11
Andrea	Moderate	1	0	1	1	1	4
Betsy	High	0	3	3	9	4	19

The children differ also in the judged function of inner speech instructions. For example, 18 of Betsy's 19 inner speech instances (95%) clearly showed planning. Only six of Mike's eleven inner speech instances (56%) appeared to serve planning. Andrea's four instances favored planning by a three-to-one ratio. Thus while the amount of inner speech did not seem to be directly related to emergent reading ability, occurrence within the two modes and clarity with which the speech served planning seems to be related to level of emergent reading.

DISCUSSION

It should be recalled that this paper reviews data from a small sample of children at a specific site and the methodology and analyses are exploratory. Therefore, the results are only suggestive. These suggestions can be compared with other theoretical and empirical sources investigating the various roles of dialogue and monologue with young children (Blank, 1973; Bruner, 1975; Vygotsky, 1962; Wells, 1981).

The results of the effectiveness rankings indicate that both naive and experienced linguistic judges, given a typed transcript of the samples, can evaluate the abilities of children to produce effective monologue and dialogue. Experienced judges were also able to agree about intrusions of inner speech into each mode.

Generally, Mike's and Betsy's relative standings would support the claim that dialogue competence precedes monologue competence. Differences in the modes into which inner speech intruded for these children also supported the dialogue before monologue expectation.

Andrea, however, seems to exhibit a different developmental pattern. In her case, monologue competence is developing well when compared to her dialogue skills, as exhibited in school. Throughout the kindergarten year, Andrea was seen as an unusually "shy" child. She could, but rarely did, take active, elaborated part in conversation but she could gain attention and aid in non-verbal turn-taking activities. While Mike seemed not able to take responsibility for maintaining dialogue, Andrea seemed able but not willing.

We have speculated that the amount of control over the planning function of inner speech may be related to developing literacy. For example, Mike who is still at a very early level of emergent reading ability used five times as many instances of inner speech that could not be classified as planning as did the other children. In addition, Mike often complained to the first author of having ideas in his head that he was trying to get out, as if control of his thoughts was not yet fully attainable. The two girls seemed to have a great deal of control. Betsy could even explain to an examiner why her wording was more precise than alternate wordings.

The total transcripts reveal some support for the Vygotskyian notion that the planning function of inner speech develops in dialogue between individuals. For example, in Mike's monologue productions, he frequently employed inner speech only after it was modelled by the examiner through dialogue that served as metacognitive guidance (Wertsch, 1979).

In one instance, Mike was asked to tell a make-believe story about Little Prince Charming learning to ride a big wheel. Mike had shown the requisite knowledge in an earlier conversation with the examiner; however when asked to tell the entire story "from beginning to end," Mike contributed story elements only in response to adult interrogatives such as, "Where was he?" "Who was with him?" When asked to retell the story himself, Mike again managed to elicit cues from the adult. Finally, when asked to dictate the story, Mike inserted at least two of the adult's planning cues (underlined) into his dictation:

He did it on Mars.
And where it happened.
And rode one.

In the early sessions of the study, Mike would make no monologue attempts without dialogic exchanges. His behavior was consistent with the lowest levels of a scale for judging emergent reading ability (Sulzby, 1982; Otto & Sulzby, this volume). By late spring, Mike began to question whether planning insertions, such as, "And where it happened," were actually part of his story. The inference we draw is that, earlier, Mike was employing dialogically-imparted planning structures from the adult for a process that he had not yet internalized. As a result, he did not distinguish these structures from his own content. He did not seem to be able to use them independently as covert inner speech guiding his composition nor to insert them in changed form into his dictation. By spring, however, recognition of their inappropriateness in externalized language was emerging.

In summary, we suggest that examination of dialogue and monologue performance by young children is possible and is potentially informative about their emerging knowledges about literacy. Qualitative analysis suggests that, for some children, intrusions of inner speech can be informative signs of the status of internalized speech forms. These findings are only exploratory and must be combined with other indicators of emerging literacy.

NOTES

1. Sulzby distinguishes between overall emergent reading abilities and emergent abilities for specific literacy tasks. For example, she has developed one scale for assessing emergent reading abilities for dictated and handwritten stories and a second scale for assessing emerging abilities to read favorite storybooks.
2. "Planning" examples could be judged by relationship to intended goals. The examples called "other" often lacked cues tying them to goals. We first characterized these instances as "unloadings," appearing similar to an externalized stream of consciousness. Sometimes these "other" statements are clarified by the context as in this example by Mike. He was asked what he thought another child would do with a story Mike had written for him. Mike answered, "This is chapter one, this is chapter two, this is chapter three." It was not until the examiner asked the question again that he elaborated, "Maybe read it or something," a possible explanation of his original statement.

ACKNOWLEDGEMENTS

The research reported herein was supported in part by a grant from the National Institute of Education, NIE-G-80-0176, to the second author. Statements made are, however, solely the responsibility of the authors. We thank the children and their teacher who must remain anonymous. Additionally, we thank June Barnhart, Clifford Cox, Paula Olszewski, and Beverly Otto for their help in scoring data and reacting to the paper.

REFERENCES

Blank, M. *Teaching learning in the preschool: A dialogue approach*. Columbus, Ohio: C. E. Merrill, 1973.

Bruner, J. S. Language as an instrument of thought. In A. Davies (Ed.), *Problems of language and learning*. London: Heinemann Educational Books, 1975.

Cherry, L. J. The role of adults' requests for clarification in the language development of children. In R. O. Freedle (Ed.), *New directions in discourse processing*. Vol. II. Norwood, N.J.: Ablex, 1979.

Otto, B., & Sulzby, E. Judging the emergent reading abilities of kindergarten children. *Thirty-first National Reading Conference Yearbook* (this volume).

Sulzby, E. *Beginning readers' developing knowledges about written language*. Proposal to the National Institute of Education. Evanston, Ill.: Northwestern University, 1980.

Sulzby, E. *Kindergarteners begin to read their own compositions: Beginning readers' developing knowledges about written language project*. Final report to the Research Foundation of the National Council of Teachers of English. Evanston, Ill.: Northwestern University, 1981.

Sulzby, E. Oral and written language mode adaptations in stories by kindergarten children. *Journal of Reading Behavior*, 1982, *14*, 51–59.

Vygotsky, L. S. *Thought and language*. Cambridge, Mass.: MIT Press, 1962.

Wells, G. *Learning through interaction: The study of language development*. Cambridge: Cambridge University Press, 1981.

Wertsch, J. V. From social interaction to higher psychological processes: A clarification and application of Vygotsky's theory. *Human Development*, 1979, *22*, 1–22.

LINDA B. GAMBRELL
University of Maryland

INDUCED MENTAL IMAGERY AND THE TEXT PREDICTION PERFORMANCE OF FIRST AND THIRD GRADERS

It is generally agreed that mental imagery has a complex relationship to cognition and language (Hebb, 1966; Piaget & Inhelder, 1971). The positive effects of induced mental imagery have been carefully documented for paired associate learning (Paivio & Yuille, 1967; Wolff & Levin, 1972; Pressley; 1976) and for listening comprehension (Shimron, 1975; Guttman, Levin & Pressley, 1977). While recent research has suggested that induced mental imagery facilitates reading comprehension (Pressley, 1976; Steingart & Glock, 1979; Linden & Wittrock, 1981) the value of using imagery as a strategy for enhancing reading comprehension has yet to be thoroughly explored (Tierney & Cunningham, in press).

The primary hypothesis set forth in this study was that mental imagery facilitates access to prior knowledge and therefore enhances the ability to infer and make text relevant predictions. The ability to anticipate and make predictions based upon text information appears to be associated with efficient learning (Smith, 1978). Good readers predict more frequently (Olshavsky, 1976–77) and more accurately than poor readers (Olshavsky & Kletzing, 1978). The research of Kail, Chi, Ingraham and Danner (1977) strongly suggests that the ability to infer and predict improves with age and is dependent upon the ability to access and manipulate information from memory. In the present study induced mental imagery was investigated as a strategy for accessing prior knowledge in order to facilitate text relevant predictions.

METHOD

Subjects

Twenty-nine first graders and 29 third graders enrolled in two suburban Maryland public schools served as subjects in the study. The first graders were judged to be reading proficiently at the first grade level by their classroom teachers. Metropolitan Reading Readiness Scores ranged from 44 to 99 with a mean of 73.509. The criteria for inclusion in the study for third graders were: (1) reading comprehension scores obtained from the California Achievement Test of at least 2.0 grade level, (2) Cognitive Abilities Test scores between the 15th and 90th percentile, and (3) teacher verification of reading ability at or above the 2nd grade level.

Materials

The stimulus materials used in this study consisted of two versions of a short story, one written at the first grade readability level and one written at the third grade readability level (Fry, 1977). The first grade story (approximately 200 words) and the third grade story (approximately 235 words) contained the same number of major propositions. The events and story information were consistent across the two versions of the story.

The story was divided into five sections based upon prediction points (logical points in the story where predictions about forthcoming events could be made). Following each section of the story a prediction question was posed: "What do you think is going to happen next in the story?" One probe, "Anything else?" was allowed.

Procedure

Subjects were randomly assigned, by grade level, to either the experimental or control group. The subjects met individually with the investigator for approximately 30 minutes. All subjects were told that they would be reading a story and answering some questions about an animal called a "monfur". In the experimental group subjects were given instructions to "Make pictures in your head about what you read to help you remember . . . " prior to reading each of the five sections of the story. In the control group subjects were given instructions to "Think about what you read to help you remember . . . ".

Immediately following the silent reading of each of the five sections of the story subjects were asked, "What do you think is going to happen next?" All responses were tape recorded. All subjects read the entire story and responded to a total of five prediction questions.

RESULTS

Responses to the prediction questions were transcribed for analysis. Two raters scored the responses for thirty percent of the subjects which resulted in an interrater reliability of .90. One rater scored the remainder of the data. The raters were unaware of whether the responses were made by subjects from the experimental or control group.

The responses to the prediction questions were scored for: (1) reference to facts which were explicitly stated in the text, (2) prediction statements, and (3) accuracy of prediction statements. Data on the number of facts and predictions for first and third grade subjects were analyzed using analysis of variance procedures. Only three instances of inappropriate predictions were noted, all occurring at the first grade level, therefore, accuracy of prediction was not included in the final analysis.

Mean number of facts and mean number of prediction statements constituted the dependent variables in this study. Table 1 presents the means for the first grade subjects and Table 2 presents the means for third grade subjects in the imagery and no imagery groups. Preliminary inspection of the data revealed no

TABLE 1
Means and Standard Deviations for First Graders

	Imagery (N = 16)		**No Imagery (N = 13)**	
	$\bar{M}$	S.D.	$\bar{M}$	S.D.
Facts	6.500	2.921	5.307	3.172
Predictions	4.000	3.204	3.384	3.150

TABLE 2
Means and Standard Deviations for Third Graders

	Imagery (N = 14)		No Imagery (N = 15)	
	$\bar{M}$	S.D.	$\bar{M}$	S.D.
Facts*	12.714	7.021	6.533	3.398
Predictions*	9.928	6.293	4.533	2.748

*significant at the .005 level of significance

appreciable sex differences for first and third grade subjects. Data were combined for males and females in the subsequent analysis.

In the analysis for variance no statistically significant differences were found for first grade subjects on number of facts, $F(1,27)=1.11$, $p>.05$, or number of predictions, $F(1,27)=.27$, $p>.05$. For third grade subjects statistically significant differences in favor of the induced mental imagery group were found on both number of facts, $F(1,27)=9.31$, $p<.005$, and number of predictions, $F(1,27)=9.17$, $p<.005$.

DISCUSSION

The findings of this study provide support for Pressley's (1976) conclusion that induced mental imagery facilitates reading comprehension for children of third grade and above. Pressley's (1976) results, and the results of this study, indicate that induced mental imagery does not enhance reading comprehension for beginning readers. We know, however, from the work of Kosslyn (1976) and Levin and Divine-Hawkins (1974) that induced mental imagery appears to facilitate *listening* comprehension for children as young as first grade. Kosslyn's (1976) research further suggests that younger children rely more on imagery in memory retrieval of orally presented information than do older children.

Several rival hypotheses can be entertained to explain the discrepancy between the effects for induced mental imagery for first and third graders in this study. The findings may reflect an age/experience effect with respect to mental imagery under reading conditions. Another, perhaps more obvious explanation, is that proficient readers benefit from induced mental imagery while, for younger, less proficient readers inducing mental imagery adds an additional processing burden (processing mental images and print simultaneously). However, an inspection of the means for the first graders in the imagery and non-imagery groups suggests that there were no detrimental effects from instructions to induce mental imagery.

The results of this study support the hypothesis that induced mental imagery facilitates access to prior knowledge and therefore enhances the ability to infer and make text relevant predictions for third graders. This hypothesis was based on three assumptions: first, that the use of mental imagery facilitates interaction between the reader and the text; second, that the use of mental imagery facilitates access to text relevant prior knowledge; third, that induced mental imagery is more effective than spontaneous imagery.

The magnitude of the effect of induced mental imagery upon the recall of facts stated in the story and the generation of text relevant predictions for third grade subjects in this study is of particular interest. These findings strongly suggest that reader-text interaction is enhanced by induced mental imagery. Other studies investigating the effects of mental imagery on reading comprehension have generally reported statistical differences of such low magnitude that they are of questionable educational value (Pressley, 1976). Previous research on reading and mental imagery has focused on subject's responses to literal and, to a lesser extent, inferential questions. In this study subjects responded to open-ended prediction questions. While the speculation is tentative, it may be that imagery facilitates language elaboration and that open-ended prediction questions are more sensitive to the effects of induced mental imagery than more focused cued recall questions.

In this study the findings for third grade subjects suggest that they were highly sensitive to instructions to induce mental imagery about what they read. While there is evidence to suggest that some children spontaneously image (Kosslyn, 1976; Olshavsky, 1976–77) the results of this study support the contention that induced mental imagery is more effective than spontaneous imagery for third graders.

Clearly, a great deal of additional research must be done before we can hope to adequately understand the role of induced mental imagery in the reading comprehension process. However, it does appear that Kosslyn's (1976) hypothesis that young children use mental imagery to access semantic information in listening situations can be extended to reading situations for third graders. The third grade subjects in this study showed a remarkable ability to infer and make text relevant predictions when given instructions to induce mental imagery.

ACKNOWLEDGEMENTS

The research reported in this paper was supported by a General Research Board Award from the University of Maryland. The author is particularly grateful to Renee Brimfield, Vivian Belcher, and Ruby Bales for their participation in piloting procedures and analysis of transcripts.

REFERENCES

Guttmann, J., Levin, J. R., & Pressley, G. M. Pictures, partial pictures, and young children's oral prose learning. *Journal of Educational Psychology*, 1977, *69*, 473–80.

Hebb, D. O. *A textbook of psychology*. Philadelphia: Saunders, 1966.

Kail, R., Chi, M. T., Ingraham, A. L., & Danner, F. Constructive aspects of children's reading comprehension. *Child Development*, 1977, *48*, 684–88.

Kosslyn, S. M. Using imagery to retrieve semantic information: A developmental study. *Child Development*, 1976, *47*, 434–44.

Levin, J. R., Divine-Hawkins, P. Visual imagery as a prose learning process. *Journal of Reading Behavior*, 1974, *6*, 23–30.

Linden, M., & Wittrock, M. C. The teaching of reading comprehension according to the model of generative learning. *Reading Research Quarterly*, 1981, *18*, 44–57.

Olshavsky, J. E. Reading as problem solving: An investigation of strategies. *Reading Research Quarterly*, 1976–77, *12*, 654–74.

Olshavsky, J. E., & Kletzing, K. Prediction: One strategy for reading success. Paper presented at IRA, Houston, 1978 (ERIC Document Reproduction Service No. ED 158 264).

Paivio, A., & Yuille, J. C. Mediation instructions and word attributes in paired-associate learning. *Psychonomic Science*, 1967, *8*, 65–66.

Piaget, J., & Inhelder, B. *Mental imagery in the child*. New York: Basic Books, 1971.

Pressley, G. M. Mental imagery helps eight-year olds remember what they read. *Journal of Educational Psychology*, 1976, *68*, 355–59.

Shimron, J. Imagery and comprehension of prose by elementary school children. (Doctoral dissertation, University of Pittsburgh, 1974.) *Dissertation Abstracts International*, 1975, *36*, 795-A, (University Microfilms No. 75–18, 254).

Smith, F. *Understanding reading*. New York: Holt, Rinehart and Winston, 1978.

Steingart, S. R., & Glock, M. D. Imagery and the recall of connected discourse. *Reading Research Quarterly*, 1979, *15*, 66–83.

Tierney, R. J., & Cunningham, J. W. Research on teaching reading. In P. D. Pearson (Ed.), *Handbook on research in reading*. New York: Longman, in press.

PAULA C. GRINNELL
The University of Texas at Dallas

READING COMPREHENSION AND PICTURE USAGE: A STUDY WITH FIRST-GRADERS

In recent years there has been a controversy over the value of using pictures to teach word recognition and reading comprehension to children in kindergarten through third grade. Most of the research has been done on word recognition and indicates that pictures assist word learning but not necessarily retention of isolated words (e.g., Wood, 1976).

The data on the contribution of pictures to the reading comprehension of young children are sparse and inconclusive. Only six studies have been reported. One found that pictures interfered with poor readers' comprehension (Weintraub, 1960); three showed no significant differences (Koenke, 1968; Koenke & Otto, 1969; Miller, 1938); one discovered that good readers' comprehension improved when a picture was present (Bluth, 1972); and one showed that poor readers remembered significantly more about illustrated characters than about unillustrated characters (Waldner, 1978).

The purpose of this study was to measure first-graders' explicit comprehension of a basal reader story under three conditions: (a) reading the text alone, (b) reading the text accompanied by supporting pictures, or (c) reading the text accompanied by contradictory pictures. The use of pictures that contradicted the text was a novel aspect of this study. If any pictures can interfere with comprehension, as for instance Samuels' (1967) focal attention hypothesis suggests they do for word recognition, then contradictory pictures accompanying text would be a likely condition in which a distracting effect could be observed. In addition to the quantitative study, a picture usage instrument was developed to analyze whether the subjects used pictures and the purposes for which they used them. This was particularly important since in previous research on reading comprehension no attention was paid to whether subjects were using pictures or the functions for which they used the pictures.

METHOD

Materials

Three versions of a basal reader story, "The Little Knight" (Rowe, 1971), were prepared as follows: (a) only the text from the story was xeroxed (text alone); (b) the story was color xeroxed exactly as it appears in the basal reader (text with supporting pictures); and (c) the text from the story was color xeroxed accompanied with pictures from "Slim, Shorty, and the Mules" (Christopher, 1971) in place of the original pictures (text with contradictory pictures). The contradictory pictures were similar to the supporting pictures in terms of artist, style, color, and size. Each version of the basal reader story was assembled with a cover to simulate an actual book.

A picture usage instrument was developed to analyze the subject's predispositions toward the function and usage of pictures. The subjects were observed while they read a paragraph accompanied by a supporting picture. Then they were asked questions relating to the specific paragraph as well as more broadly based retrospective questions about the usage of pictures. (For a detailed description of the picture usage instrument, see Grinnell, 1981.)

Sample and Procedures

Forty-five suburban first-graders read aloud and retold one of the three versions of the story described above. Two weeks later, the subjects participated in an individual session using the picture usage instrument.

Scoring the Data

The story was analyzed into explicit idea units using a modification of Meyer's (1975) semantic analysis technique. Taylor's (1979) explicit recall scoring procedures were adapted and expanded (Grinnell, 1981). In addition, a method for scoring the picture usage instrument was established (Grinnell, 1981).

Three advanced graduate students in reading were trained as raters to establish the reliability of the scoring procedures. For each comparison, two raters (the researcher and another rater) worked independently, compared results, and then mediated differences (final agreement). Inter-rater reliability at final agreement was 100% for the story analysis, 99% for the retellings, and 100% for the picture usage instrument.

Data Analysis

The data analysis involved parametric, nonparametric, and descriptive procedures. To measure differences among the treatments on the number of explicit idea units stated during recall, both parametric and nonparametric statistical techniques were employed. This was done because the scatter of the data was so great that two assumptions of the parametric technique were violated. (An in-depth discussion of the rationale for statistically testing these assumptions and the results of the testing can be found in Grinnell, 1981.) In order to validly use the parametric analysis, the data points that were plus or minus three standard deviations from the mean were identified. Only one was found, and it was removed from the text alone condition (c.f. Draper & Smith, 1966).

Descriptive procedures were used to analyze the picture usage instrument. This technique was chosen because of the subjective nature of the picture usage instrument.

RESULTS

Explicit Recall

The medians for the free explicit recall data were: text alone, 34; text with supporting pictures, 48; and text with contradictory pictures, 24. The Kruskal-Wallis *H*-test on these medians demonstrated significance ($p<.02$). Nonparametric multiple comparisons indicated that subjects reading either text alone or text with supporting pictures had significantly higher scores ($p<.05$) than subjects reading text with contradictory pictures. There were no significant differences between subjects reading text alone and text with supporting pictures.

Parametric analysis of the data confirmed most of these results. The one-way analysis of variance on the explicit recall means also demonstrated significance for the free recall data ($p<.01$). The Fisher significant difference test showed that subjects reading text with supporting pictures had significantly higher scores ($p<.01$) than subjects reading text with contradictory pictures. The means and standard deviations for the recall data were: text alone, 37±26; text with supporting pictures, 49±30; and text with contradictory pictures, 24±19.

Picture Usage

The data from the picture usage instrument were analyzed to determine: (a) the functions for which the subjects used pictures; (b) the percentage of subjects who used pictures for each of the functions; and (c) the percentage of times that each of the functions was used. The following picture functions were identified by the subjects: word recognition, word recognition/comprehension, comprehension, and enjoyment. The category word recognition/comprehension was used when comments were not clearly word recognition or comprehension, e.g., "Pictures help me read."

The analysis of the percentage of subjects who used pictures showed that all of the subjects used pictures. They were distributed as follows: 30% of the subjects used pictures for word recognition, 30% used them for word recognition/comprehension, 21% used them for enjoyment, and 8% used them for comprehension. The analysis of the percentage of times that each of the functions was used showed that pictures were used 54% of the time for word recognition, 25% of the time for word recognition/comprehension, 11% of the time for enjoyment, and 5% of the time for comprehension.

DISCUSSION

The results of this study suggest the following conclusions: (a) the presence of supporting pictures with the text does not assist or hinder explicit comprehension; (b) the presence of contradictory pictures with the text hinders explicit comprehension; and (c) pictures are employed mainly to aid in word recognition. It should be noted that this study was carried out with a basal reader story. Similar research with other types of stories should be done to determine the generalizability of these results.

The fact that differential effects were found for supporting pictures and contradictory pictures has important implications regarding Samuels' (1967) focal attention hypothesis as it applies to reading comprehension. It appears that pictures can hinder reading comprehension when they contain information that is contradictory to the textual information. On the other hand, when the pictures support the textual information, they do not inhibit comprehension. Thus when determining the possibility of a distracting effect of pictures, the relationship of the pictures to the text must be taken into consideration.

The finding that related pictures do not assist in explicit reading comprehension requires further explanation. Since subjects almost never verbalized a clear awareness that pictures can function to assist comprehension, it is possible that children cannot use pictures to aid in comprehension. This may be a maturational effect (e.g., Higgins, 1978). Alternatively, children may be capable of using pictures for comprehension but have not been taught to do so. One child, when asked "Do you ever look at the pictures that go with a story?", answered "No, that would be cheating." One of the children's teachers told the researcher that "The children shouldn't be looking at the pictures. They should be reading." Furthermore, in response to the question "When you are reading a story, does your teacher ever tell you to look at the pictures in the story?", 50% of the children answered "yes", but the other 50% said "no". Thus, although children seem to use illustrations to assist in word recognition, they do not use them as readily for comprehending the story. It may be that more emphasis should be put on helping children learn how to use pictures to get the meaning of the story. To this end, children could be taught to "read" the pictures that are found in basal readers.

REFERENCES

Bluth, L. F. A comparison of the reading comprehension of good and poor readers in the second grade with and without illustration (Doctoral dissertation, University of Illinois, 1972). *Dissertation Abstracts International*, 1973, *34*, 637A. (University Microfilms No. 73-17, 122)

Christopher, S., illustrated by Rod Ruth. Slim, Shorty, and the Mules. In *Scott Foresman Reading Systems Level 4, Book G*. Glenview, Illinois: Scott, Foresman & Company, 1971.

Draper, N. R., & Smith, H. *Applied regression analysis*. New York: John Wiley & Sons, Inc., 1966.

Grinnell, P. C. The effects of pictures on first-graders' comprehension of written discourse (Doctoral dissertation, Texas Woman's University, 1980). *Dissertation Abstracts International*, 1981, *41*(10), 4266A-4267A. (University Microfilms No. 8104339)

Higgins, L. C. A factor analytic study of children's picture interpretation behavior. *Educational Communication and Technology*, 1978, *26*, 215-232.

Koenke, K. R. *The effects of a content relevant picture on the comprehension of the main idea of a paragraph* (Tech. Rep. No. 56). Madison: University of Wisconsin, Wisconsin Research and Development Center for Cognitive Learning, 1968. (ERIC Document Reproduction Service No. ED 024 540)

Koenke, K., & Otto, W. Contribution of pictures to children's comprehension of the main idea in reading. *Psychology in the Schools*, 1969, *6*, 298-302.

Meyer, B. J. F. *The organization of prose and its effects on memory*. In T. A. Van Dijk & W. O. Hendricks (Eds.), *North-Holland Studies in Theoretical Poetics* (Vol. 1). New York: American Elsevier Publishing Company, Inc., 1975.

Miller, W. A. Reading with and without pictures. *Elementary School Journal*, 1938, *38*, 676-682.

Rowe, A., illustrated by Rod Ruth. The little knight. In *Scott Foresman Reading Systems Level 4, Book F*. Glenview, Illinois: Scott, Foresman & Company, 1971.

Samuels, S. J. Attentional process in reading: The effect of pictures on the acquisition of reading responses. *Journal of Educational Psychology*, 1967, *58*, 337-342.

Taylor, B. A comparison of good and poor readers' ability to read for meaning (Doctoral dissertation, Virginia Polytechnic Institute and State University, 1978). *Dissertation Abstracts International*, 1979, *39*(4), 2108A. (University Microfilms No. 78-18560)

Waldner, D. N. The effect of pictures on the story comprehension and retention of good and poor readers (Doctoral dissertation, University of Denver, 1978). *Dissertation Abstracts International*, 1979, *39*(11), 5620B. (University Microfilms No. 79-10670)

Weintraub, S. The effect of pictures on the comprehension of a second grade basal reader (Doctoral dissertation, University of Illinois, 1960). *Dissertation Abstracts*, 1960, *21*, 1428A-1429A. (University Microfilms No. 60-4016, 92)

Wood, M. W. A multivariate analysis of beginning readers' recognition of taught words in four contextual settings (Doctoral dissertation, Texas Woman's University, 1976). *Dissertation Abstracts International*, 1977, *37*(3), 4122A. (University Microfilms No. 77-766)

JAMES V. HOFFMAN
LESA A. KASTLER
MARCIA F. NASH
JOHN DALY
University of Texas at Austin

STUDENTS' BELIEFS AND ATTITUDES ABOUT ORAL READING INSTRUCTION

The child learning to read in the context of the classroom approaches instructional tasks with a complex set of beliefs and attitudes. According to Brophy and Evertson (1981), beliefs are statements about something or someone thought to be true whether or not they are. Attitudes are affective or emotional reactions to events or people. There is a substantial—though not particularly enlightening—history of research into attitudes and reading. Most of the research in this area has focused on attitude formation and its relationship to success in learning to read. More recently, there has emerged a line of research exploring children's beliefs or conceptions of reading at various stages of proficiency (Downing, 1969; Johns, 1970). The results of these studies seem to indicate distinct patterns of belief systems which are in part developmental and in part a reflection of the environmental influences of home and school.

The study to be reported in this paper examines student beliefs and attitudes toward oral reading in a specific instructional context—teacher guided oral reading. It was hoped that the insights gained from such an investigation might ultimately lead to more enlightened classroom practices. Teacher guided oral reading has been and remains a common part of the classroom routine at primary levels (Austin & Morrison, 1967; Howlett & Weintraub, 1980). Teacher guided oral reading typically takes the form of "round-robin" or "barbershop" reading. The practice has been indicted from many sides and for many reasons (Artley, 1972; Spache & Spache, 1977). One of the most common charges is that it is unfair to students—in particular low ability readers—exposing them to ridicule and embarrassment. The results of a survey by Daly and Hoffman (1981) of classroom practices and teacher attitudes toward guided oral reading would seem to indicate that teachers do not share this belief. Most primary teachers regarded guided oral reading as a valuable instructional activity for both good *and* poor readers. They do not view the task as unfair or embarrassing to poor readers. They do, however, perceive the activity as boring for both the teacher and the students. Teachers, by and large, seem to take the position that it doesn't have to be fun to be good for you.

METHOD

The research site was a moderate sized school district located in the south central region of the United States. The developmental reading program in this system could best be described as traditional with a basal orientation and an emphasis on ability grouped instruction. All teachers of students in the classrooms studied reported frequent use of group guided oral reading as part of their instructional program. The data to be reported were collected as one part of a comprehensive investigation of teacher-pupil interactions during oral reading.

Subjects

Students from the high and low reading groups (N = 207) in 23 second grade classrooms in this district's ten elementary schools participated in the study. There were a comparable number of males and females in the sample.

Procedures

The data for this project were collected during weekly visits to the research site over a three month period in the fall of the year. The students in each group were selected at random. All tests and interviews were conducted individually outside of the classroom setting. Students were first administered the Slosson Oral Reading Test (SORT) to provide an estimate of reading achievement levels. This brief testing period was followed by an extensive interview session.

Instrumentation

The interview instrument consisted of 80 questions presented orally to students in a closed response format. Some of the items required the student to respond with a "yes" or "no" (e.g., "Do you enjoy reading?). Others required the students to state a preference (e.g., "Would you rather read out loud or read silently?"). No item contained more than two choices from which the students were to select. Six of the items were repeated exactly in other parts of the interview to check for response consistency. The items were clustered into six major sections related to the following themes: I. Attitudes; II. Perceived Ability; III. Proficiency Constructs: IV. Teacher's Role; V. Social/Evaluation Context; and VI. Models.

The items within each section had been developed and pilot tested as part of an earlier study (Hoffman, Kastler, and Nash, 1981). The first set of ten items in the Attitudes section explored students' feelings about reading in general, and the second set of ten items examined feelings toward oral reading in particular. The Perceived Abilities section had a similar breakdown with nine items focused on their silent reading ability and nine items on their oral reading ability. In the Proficiency Construct section there were five items designed to explore what the respondent knew about good oral reading performance and another five items over what the respondents knew about poor oral reading performance. The Teacher's Role section contained five items covering what the students liked the teacher to do when they made a mistake or otherwise encountered difficulty in oral reading. The Social/Evaluation Context section contained fifteen items related to how the respondents felt about others observing or judging their oral reading performance. The final section of the instrument contained five questions related to what models of oral reading the respondents are exposed to both in and out of the school setting.

RESULTS

A preliminary analysis was conducted to test the reliability of student response patterns. Tetrachoric correlation coefficients were computed for the six pairs of repeated items in the interview. The average correlation for all six pairs of items was .79. Each of the individual correlations was statistically significant at the $p < .001$ level.

Correlation matrices for each of the six major sections of the interview instrument were then computed as a test of construct validity. Separate matrices were formed for the general vs. oral reading attitudes section and the silent vs. oral reading perceived abilities section. The most highly correlated items within each section were then selected as the basis for computing a composite score on each subsection. There were no significant correlations between any of the items in the proficiency construct section so no composite score was created for this section. The Social/Evaluation Context section was broken down based on item content and inter-item correlation patterns into two new areas of audience effects and negative affect toward oral reading performance. The following areas were thus identified: (1) General Attitude (3 questions); (2) Oral Reading Attitude (5 questions); (3) Perceived Ability—silent reading (4 questions); (4) Perceived Ability—oral reading (4 questions); (5) Teacher's Role (3 questions); (6) Audience Effects (3 questions); (7) Negative Affects (3 questions); and (8) Home Reading Models (2 questions).

A multiple regression analysis was then performed using the SORT achievement test scores as the criterion variable and the eight composite scores as the predictors. The multiple R was found to be .37 ($p<.001$). This figure meets Cohen's (1977) criteria for a moderate effect size. Three of the eight composite scores were found to explain most of the variance. These were: perceived ability in oral reading; teacher role; and audience effects. A reduced multiple regression using just these three predictor variables yielded a multiple R of .34 with an adjusted R^2 of .10. The Beta weights for the three composite variables were .22 for perceived oral reading ability ($p<.002$); .20 for teacher role ($p<.004$); and .13 for enjoyment/audience effects ($p<.07$). The questions subsumed in each of these three composite variables are presented in Figure 1.

A subsample of high and low ability readers was identified next for purposes of performing an item analysis comparison of response patterns. The high ability reader group ($N=77$) consisted of readers assigned to a high reading group in their classrooms and scoring higher than the 3.5 grade level on the SORT. The low ability reader group ($N=50$) consisted of readers assigned to a low

Perceived Ability in Oral Reading

Are you a good oral reader?

Do you read out loud very well?

Do other kids think you are a good oral reader?

Does your teacher think you are a good oral reader?

Teacher's Role

Do you like for the teacher to tell you words when you don't know them?

Do you like for the teacher to call on other kids to help you with the words?

Do you like for the teacher to help you figure out words you don't know?

Audience Effects

Do you like to read out loud when the teacher calls on you in the reading group?

Do you like to read out loud to the whole class?

Do you like to read out loud to your teacher when you are alone together?

FIGURE 1

153

reading group in their classroom and scoring less than the 2.5 grade level on the SORT. Only the most striking points of contrast in response patterns will be presented in this summary.

The responses of students in both the high and low groups reflected positive feelings about reading in general. Questions which made either direct or indirect comparisons of silent and oral reading tasks revealed that both groups hold a more positive view of silent reading than oral reading. A majority of high readers (70%) reported that reading out loud is fun, while the majority of low readers (52%) reported that it was not fun. Both groups responded overwhelmingly (95% to 5%) that their teachers thought of them as good silent readers. The figure remained almost the same (93% yes) for the high group when asked whether their teacher thought of them as good oral readers. In the low group, though, the percent of students who reported that their teacher did not think of them as good oral readers rose to 25%. Only half of the low group regarded themselves as good oral readers whereas 83% of the low group regarded themselves as good silent readers. For the high group, 87% regarded themselves as good oral readers and 95% as good silent readers.

Both groups overwhelmingly preferred that the teacher help them figure out unknown words over giving them the word (89% vs. 11%). When asked whether they liked for the teacher to call on other kids to help with words, 73% of the low group regarded the practice favorably as compared to only 56% of the high group.

A vast majority (90%) of the high group readers reported that they like to read out loud in their reading group. 44% of the low group readers responded that they did not like to read out loud in their groups. This difference is in contrast to the congruent pattern of responses to the question of whether they enjoyed reading out loud to the teacher when they were alone (83% *yes* for the high group and 74% *yes* for the low group). Students in both groups revealed sensitivity to the evaluative aspects of reading orally in groups. They agreed that oral reading performance affected placement in high or low reading groups. Most students in both groups reported that they tried hard not to make mistakes when reading orally. The majority of students in both groups also reported feeling nervous when reading orally with others listening.

DISCUSSION

The findings from this study suggest that even by the beginning of second grade students have some developing beliefs and attitudes toward oral reading instruction. These beliefs and attitudes are clearly tied to reading ability. The three composite variables (i.e., perceived ability in oral reading, teacher's role, and audience effects) identified through the multiple regression analysis point to those areas where beliefs and attitudes are strongest. The better the reader the greater the enjoyment regardless of the social or performance context. Also, the better the reader the greater the desire for the teacher to assume a low profile in helping when difficult words are encountered. The poorer the reader the less the enjoyment and the greater the desire for teacher involvement. It is also interesting that the variable perceived ability in oral reading relates (i.e., predicts) reading achievement much better than does perceived ability in silent reading. Valid or not, oral reading performance seems to be the best gauge for students to use in evaluating their own ability.

The analysis of specific items relative to extreme ability levels adds additional evidence to suggest that, at least for poor readers, oral reading is a stressful and anxiety producing part of the classroom instructional routine.

ACKNOWLEDGEMENTS

This research was supported in part by a grant from the National Institute of Education (Contract No. 6800-033). It is part of an ongoing study into teacher verbal feedback to student miscues.

REFERENCES

Artley, A. S. Oral Reading as a Communication Process. *The Reading Teacher* 36 (October 1972): 46–51.

Austin, M. C., and Morrison, C. *The First R: The Harvard Report on Reading in Elementary Schools*. New York: Macmillan Publishing Co., 1963.

Brophy, J. E., and Evertson, C. M. *Student Characteristics and Teaching*. New York: Longman, Inc., 1981.

Cohen, J. *Statistical Power Analyses for the Behavioral Sciences*. New York: Academic Press, 1977.

Daly, J. A., and Hoffman, J. V. *Teacher Attitudes toward Oral Reading Instruction*. Unpublished manuscript, The University of Texas at Austin, 1981.

Downing, J. How Children Think About Reading. *The Reading Teacher* 23 (December 1969): 217–230.

Hoffman, J. V., Kastler, L. A., and Nash, M. F. *A Study of Students' Perceptions of the Teacher's Role During Guided Oral Reading*. Paper presented at the Ninth Southwest Regional Conference of the International Reading Association, San Antonio, 1981.

Howlett, N., and Weintraub, S. Instructional Procedures. In R. Calfee and P. Drum (Eds.), *Teaching Reading in Compensatory Classes*. Newark, Delaware: International Reading Association, 1979, 87–103.

Johns, J. L. Reading: A View from the Child. *The Reading Teacher* 23 (April 1970): 647–48.

Spache, G. D., and Spache, E. B. *Reading in the Elementary School*. Boston: Allyn and Bacon, Inc., 1977.

JAMES V. HOFFMAN
SHARON O'NEAL
RICHARD O. CLEMENTS
University of Texas at Austin

THE EFFECTS OF DIFFERENTIATED PATTERNS OF VERBAL FEEDBACK TO MISCUES ON WORD IDENTIFICATION STRATEGIES AND SUCCESS

Research on teacher feedback to oral reading miscues has received considerable attention over the past few years (Niles, 1980). This increased activity can be attributed in part to the development of miscue analysis techniques and theory (Goodman, 1967; Goodman and Burke, 1972) and in part to the findings from classroom research which have documented the wide use of oral reading in primary classrooms (e.g., Hoffman & Clements, 1981). Research into teacher verbal feedback to miscues can be classified into two basic kinds of studies. The first set of studies are field-based in nature and have attempted to describe the characteristics and distribution of types of teacher feedback during oral reading instruction. A number of studies in this group have explored differences in feedback in relation to ability level of students (e.g., Allington, 1978; Hoffman and Clements, 1981; Pflaum, Pascarella, Boswick, & Aver, 1980). Others have examined the relationship between teacher beliefs about reading and their feedback patterns (Hoffman and Kugel, 1981; Mitchell, 1980). The other broad area of work in teacher feedback has been more clinical in nature and focused on comparisons of selected patterns of feedback as they relate to student performance characteristics (Jenkins and Larson, 1979; Niles, Graham, and Winstead, 1977 and 1978; Niles, 1979). While studies of this latter group point to the existence of relationships, the treatments have at times been unclearly differentiated and left many crucial features of feedback such as timing and point of interruption unaccounted for in the final analysis.

The study to be reported falls clearly into the second line of research in that it is clinical in nature. The purpose of the study was to explore the ways in which variation in the form and timing dimensions of feedback (Hoffman, 1979) relate to differences in pupil performance.

METHOD

The study was conducted in the public school system of a moderate size city in the south central area of the United States. All ten of the elementary schools in the district participated in the study.

Subjects

There were 84 students in the study who had been selected proportionally from the high reading groups of twenty second grade classrooms in the school district. Students were given the Slossen Oral Reading Test as a measure of general reading achievement. Those scoring below the 1.0 grade level on this test were excluded from consideration. These subject selection critieria were used to avoid having students participate in the experiment for whom the text materials would be too difficult.

Materials

Two 150 word passages were developed for use as experimental materials in this study. Both passages were written at approximately a second grade difficulty level. The control over passage difficulty was achieved by (1) limiting the average sentence length to around nine words; and (2) restricting the general vocabulary to words found in locally adopted first and second grade basal readers. The only exceptions to the vocabulary control restriction were two different sets of four difficult words included in each passage (Passage A: slope, vicious, echo, invade; Passage B: gaze, serious, jewel, remote). These eight difficult words were chosen from the fourth grade level vocabulary list of the Burns and Roe *Informal Reading Assessment* (1980). Each difficult word appeared twice in a passage. No two difficult words appeared in the same sentence in either passage.

Design

Six different treatment conditions were devised for use in this study. These six conditions varied across the two dimensions of form and timing of feedback. The basic feedback forms were (1) terminal-or supplying the word to the student; (2) sustaining grapho-phonic-or attempting to help the student identify the text word by focusing attention on orthographic features of the word (i.e., "Look carefully at the letters in the word" and "Try to sound it out."); and (3) sustaining context-or attempting to help the student identify the text word by focusing attention on surrounding structures and meanings (i.e., "Let's try reading that sentence again" and "Does that word make sense? What word would fit better?"). The prompts used for the sustaining conditions were developed based on high frequency strategies used by teachers in earlier field studies. The timing of feedback was varied in terms of the point of interruption: immediate (before or immediately after the word following the miscue); and delayed (at the first sentence back following the miscue).

Procedures

The students were randomly assigned to treatment conditions. They were asked to read aloud both specially designed passages under their assigned feedback condition. Feedback was offered to the students only for those miscues made on the eight difficult words. This made a total of 16 opportunities for feedback given that each of the difficult words occurred twice. Student miscues on all other words were ignored. These sessions were tape recorded. The sessions were later reviewed by the researchers for accurate implementation of the treatment condition. If upon review it was found that the experimenter failed to give the correct feedback on over 10% of the student miscues in difficult words then that subject's data was discarded. This review process resulted in the deletion of data from only two subjects.

Data Analysis

Oral reading performance on the sixteen experimental words in this study was coded using a modified version of the FORMAS taxonomy (Hoffman and Baker, 1981). The words were coded initially for miscue type (omission, substitution, mispronunciation, call for help, hesitation, and repetition). Omissions, substitutions and mispronunciations were further classified for high and low meaning change and substitutions and mispronunciations were classified for graph-phonic similarity. Each miscue was also categorized for the subject's immediate reaction to his/her miscue (continuations, repeated attempts, pauses, calls for help, self-corrections) and for the ultimate resolution of the miscue (teacher identifies miscue, student identifies miscue). The other words in the text were coded only for the occurrence of a miscue. Expert coders listened to tape recordings of the experimental sessions to code reading performance.

For each of the categories of miscues described above, the subject's errors were expressed as a percentage of the opportunities for error in that category. This was the dependent variable used in analyses described below except where otherwise noted. The basic design of this experiment included three factorialized between subject variables. These were the timing of the feedback, the form of the feedback and the reading ability of the subject. Students whose reading scores exceeded the sample mean (4.7 grade level) on the SORT were classified as higher ability readers ($\bar{x}=5.7$). Students who scored below the sample mean were classified as lower ability readers ($\bar{x}=3.3$). It should be noted that the classification by ability is a relative one.

A series of five analyses of variance were run using this design with the addition of one of the FORMAS within subject variables (miscue type, meaning change, grapho-phonic similarity, reaction, and resolution). Minor changes from analysis to analysis are discussed in the results section. In every case an unweighted means solution was used to solve the problem of unequal cell sizes.

A final analysis was run in which the percent of second miscues which were also missed the first time were analyzed as a function of the form of feedback, the timing of feedback, and reading ability.

RESULTS

Preliminary analyses revealed no statistically significant differences between the six treatment groups on either the percentage of miscues on target words or on the total number of miscues made including those outside the target words. The overall error rate was 41% on the target words and 6% when all miscues were considered. There was a statistically significant difference ($p<.001$) between the high and low ability readers in their error rates on target words (23% versus 59%) and on all text words (4% versus 8%).

The first basic set of analyses examined the percent of miscues on target words as a function of miscue type, form of feedback, timing of feedback, and ability group. For these analyses and those that follow omission, insertions, and calls for help were not considered due to their low frequency (i.e., less than 1% of the total). The distribution of miscues across the remaining three categories differed significantly across all treatment and ability groups [$F(2,144)=29.04$, $p<.01$]. The most frequent types of miscues across all groups were mispronunciations (56%) followed by hesitations (29%) and then substitutions (15%). There were no statistically significant main effect differences in the distribution of miscue types among the six treatment conditions or between ability groups. There was, however, a statistically significant four-way interaction between miscue type, feedback type, feedback timing and ability groups [$F(4,144)=531$, $p<.01$]. The data for higher and lower ability readers relative to this interaction are presented in Table 1. The major source of the interaction seems to be related to differential performance of the poorer readers under the various feedback conditions.

Subjects in the lower ability group under the delayed grapho-phonic feedback

TABLE 1
Percent of Miscues as a Function of Miscue Type, Form of Feedback, Timing, and Ability Group

	Form	Timing	Substitutions	Mispronunciations	Hesitations
Lower Ability	Terminal	Imm.	4.33	27.83	25.50
		Delay	14.38	22.06	15.06
	Sus. G.P.	Imm.	7.78	25.94	29.39
		Delay	14.40	33.20	10.60
	Sus. Con.	Imm.	8.67	46.92	4.25
		Delay	6.31	17.38	25.06
Higher Ability	Terminal	Imm.	.81	22.19	3.19
		Delay	.92	9.93	2.71
	Sus. G.P.	Imm.	8.33	12.67	7.42
		Delay	1.95	11.50	7.00
	Sus. Con.	Imm.	1.86	11.71	4.50
		Delay	.00	20.63	3.25

condition had a much higher incidence of mispronunciation type miscues as compared to those in the immediate grapho-phonic condition. Under the immediate condition the lower ability readers in the grapho-phonic treatment condition tended to hestitate rather than mispronounce. The lower ability readers in the delayed context feedback condition demonstrated a much higher incidence of hesitation type miscues as compared to those in the immediate context prompt condition. Lower ability readers in the immediate context condition tended to mispronounce rather than hesitate.

The patterns for the higher readers under the sustaining conditions tended to be in direct contrast to those of the lower readers in particular with respect to mispronunciations. Immediate grapho-phonic feedback inflated the level of mispronunciations. Immediate context diminished mispronunciations. Under the delay conditions those patterns were reversed for the higher ability readers. Terminal feedback under both immediate and delayed conditions had similar effects for both higher and lower ability readers.

Reader performance was analyzed next under sustaining feedback conditions for the two most frequent miscue types (mispronunciations and hesitations) to determine whether it was the teacher or the student who was ultimately responsible for identifying a text word once a miscue had been made. There was a statistically significant ($p < .001$) five-way interaction between miscue type, feedback type, feedback timing, ability group and resolution (Table 2). For the higher reading group, the sustaining context conditions—both immediate and delayed—were superior to any of the other conditions in eliciting student over teacher identification on mispronunciations. The same was true for the lower readers in the delayed condition. In the immediate condition, however, the context group was the highest in teacher identification of mispronunciation miscues.

The final area of analysis focused on the percent of miscues made on a target word the second time it was encountered given a miscue on the first encounter. The percent of errors were analyzed as a function of feedback type, feedback

TABLE 2
Percent of Miscues as a Function of Miscue Type, Form of Feedback, Timing, Ability Group, and Resolution

	Form	Timing	Mispronunciations		Hestitations	
			Teacher ID	Student ID	Teacher ID	Student ID
	Terminal	Imm.	15.67	12.00	21.83	3.33
		Delay	17.38	4.75	7.88	5.38
Lower Ability	Sus. G.P.	Imm.	13.78	8.44	23.00	5.56
		Delay	20.80	12.80	7.60	2.60
	Sus. Con.	Imm.	37.33	9.67	4.17	.00
		Delay	7.13	10.00	13.38	9.63
	Terminal	Imm.	16.63	5.38	2.38	.75
		Delay	5.43	4.57	2.71	.00
Higher Ability	Sus. G.P.	Imm.	7.33	5.17	3.00	4.33
		Delay	5.60	5.70	3.20	3.20
	Sus. Con.	Imm.	2.71	7.00	2.71	.00
		Delay	3.00	15.75	1.50	1.50

timing and ability group. These error rates are presented in Table 3. A statistically significant main effect was found for timing on error rate [F(1,72=5.49, p<.05]. Although not reaching levels of statistical significance, the delayed context feedback condition was superior to all other conditions in reducing the incidence of repeated errors.

TABLE 3
Percent of Repeated Errors on Second Encounter of Target Words as a Function of Feedback Form and Timing

	Form	Immediate	Delayed
	Terminal	53.33	53.54
Lower Ability	Sus. G.P.	59.67	55.33
	Sus. Con.	72.02	37.80
	Terminal	27.08	14.29
Higher Ability	Sus. G.P.	34.52	10.33
	Sus. Con.	14.76	8.33

DISCUSSION

The results of this study clearly indicate that differences in verbal feedback can affect the quality of student performance during oral reading. The precise nature of the relationship between teacher feedback and student performance is complicated but the findings of this study point toward some valuable hypotheses useful in guiding future investigations.

The tendency for the lower ability readers in the immediate grapho-phonic condition to hesitate can be explained in part as "learned helplessness." That is, the readers come to recognize that they will receive help soon if they just wait. There may also be a certain amount of error avoidance operating within this condition for the lower readers. By hesitating rather than mispronouncing they avoid having their own decoding efforts corrected immediately. In the delayed grapho-phonic condition these factors are not in force. The lower readers have time to apply the strategy communicated implicitly in the feedback and they mispronounce.

Lower ability readers in the immediate context condition respond with inflated mispronunciations over hesitations. One explanation for the willingness of this group to mispronounce may be that they recognize that with immediate feedback pending there is no time or way to utilize the context strategies implicit in the feedback. They have limited time to process and only the context up to the point of the miscue to work with in making an attempt. The option of mispronouncing is open to them because even if unsuccessful their effort will not be challenged by the context oriented prompt. The tendency for lower ability readers in the delayed context condition was to hestitate. As with the delayed grapho-phonic group it seems that these readers had sufficient time to apply the strategy implicit in the feedback offered. We hypothesize that these readers are using hesitations as thinking time or even for covert rereading. Timing seems to be the critical factor influencing the ability or the willingness of the lower readers to apply certain strategies. For the higher readers timing did not seem to be so crucial.

Delayed context shows up in two areas as a potentially valuable feedback strategy. Students reading in this condition were more successful than in any other in identifying their own miscues. Students in this group were also highly successful in identifying target words in their second encounter. Interestingly, the students in the immediate context condition were among the poorest in both these areas of performance.

The importance of timing of feedback showed its overall impact in the area identifying the target words on their second encounter. The effect of delaying feedback was salient across all forms of prompts for higher and lower readers alike.

ACKNOWLEDGEMENT

This research was supported in part by a grant from the National Institute of Education (NIE Contract No. 6–800–033).

REFERENCES

Allington, R. L. *Are good and poor readers taught differently? Is that why poor readers are poor readers?* Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March 1978.

Burns, P. C., & Roe, B. D. *Informal reading assessment*. Chicago: Rand McNally Publishers, 1980.

Goodman, K. S. Reading a psycholinguistic guessing game. *Journal of the Reading Specialist*, 1967, *4*, 126–135.

Goodman, Y., & Burke, C. *Reading miscue inventory manual: Procedure for diagosis and evaluation*. London: Collier and McMillan, 1972.

Hoffman, J. V. On providing feedback to reading miscues. *Reading World*, 1979, *18*, 342–350.

Hoffman J. V., & Baker, C. Characterizing teacher feedback to student miscues during oral reading instruction. *Reading Teacher*, 1981, *34*, 907–913.

Hoffman, J. V., & Clements, R. O. *A descriptive study of the characteristics of miscue focussed verbal interactions between teacher and students during guided oral reading*. Paper presented at the Annual Meeting of the International Reading Association, New Orleans, 1981.

Hoffman, J. V., & Kugel, C. *A study of theoretical orientation to reading and its relationship to teacher verbal feedback during reading instruction*. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, 1981.

Jenkins, J. R., & Larson, K. *Evaluating error correction procedures for oral reading* (Tech. Rep. No. 55). Champaign, IL: Center for the Study of Reading, University of Illinois, 1978.

Mitchell, K. A. Patterns of teacher-student responses to oral reading errors as related to teachers' theoretical framework. *Research in the Teaching of English*, 1980, *14*, 243–263.

Niles, J. A. *The effects of selected teacher prompting strategies on oral reading performance*. Paper presented at the Annual Meeting of the National Reading Conference, San Antonio, 1979.

Niles, J. A. *Effects of teacher feedback on beginning readers' oral reading behavior*. Paper presented at the Annual Meeting of the National Reading Conference, San Diego, 1980.

Niles, J. A., Graham, R. T., & Winstead, J. C. Effects of teachers' response strategies on children's oral reading performance: Case of interference? *Reading in Virginia*, 1977, *5*, 25–28.

Niles, J. A., Graham, R. T., & Winstead, J. C. *The effects of teacher feedback responses on children's oral reading performance*. Paper presented at the Annual Meeting of the International Reading Association, Miami, May 1978.

Pflaum, S. W., Pascarella, E. T., Boswick, W., & Aver, C. The influence of pupil behaviors and pupil factors on teacher behaviors during oral reading lessons. *Journal of Educational Research*, 1980, *74*, 99–105.

CONNIE JUEL
DIANE ROPER/SCHNEIDER
University of Texas at Austin

THE GROWTH OF LETTER-SOUND CORRESPONDENCE KNOWLEDGE IN FIRST GRADE AND ITS RELATION TO READING ACHIEVEMENT AND PROGRAMS

The primary focus of this study was to monitor the growth of letter-sound correspondence knowledge in first graders. We were interested in whether the development of letter-sound correspondence knowledge was fostered or hindered by instruction in two very different types of reading series. In addition, we wanted to see if letter-sound correspondence knowledge related to reading achievement throughout, as well as at the end of, the first grade year.

There were two basic research questions: 1) What relationship existed among instructional materials, other factors (e.g., entering Metropolitan Readiness scores) and the growth and application of letter-sound correspondence knowledge? and 2) How important is letter-sound correspondence knowledge in the acquisition and performance of reading skills in first grade?

METHOD

Sample

The sample consisted of 93 children—sixty-one Anglo and thirty-two other, most of whom were Mexican American. Approximately half the students were male and half female. The children were non-readers upon entering first grade as determined by teacher and experimenter screening on reading word lists and informal reading inventories. However, the children had no known reason why they should not succeed in learning to read (e.g., no known records of unusual kindergarten behaviors). All children were fluent English speakers, scored above the 40th percentile on the total Metropolitan Readiness Test, and were placed in the middle reading group within their classrooms.

The children came from eleven classrooms in three schools. School #1 was an "open concept" structure in a middle to upper middle class neighborhood. Its students were predominantly Anglo. School #2 was an older, more traditional structure in a middle to lower middle class neighborhood. Some of its students were bussed to achieve racial integration. School #3 was an older structure to which exterior halled classrooms had been added. It was in a lower to middle SES area and was naturally integrated. There was no significant difference between schools as shown by the Metropolitan's national percentile norms.

Approximately half the students were placed in a phonics oriented series with text and instruction which emphasized primarily regular decodable patterned words (the 1972 Economy series). The other half of the children were in a series which focused on high interest stories and whose text and instruction, therefore, had more equality between regular and irregularly decodable patterned words (the 1971 Houghton Mifflin series). Both texts were used in all three schools.

Procedure

Nine tests were given during the 1980–81 school year. The Metropolitan Readiness Test and the Iowa Test of Basic Skills were administered by the district in September and April, respectively. The Basal Core Word List Test was given in Nov./Dec., February, and May. The Bryant Test of Basic Decoding Skills was given in Nov./Dec., February, and April. This last Bryant testing coincided with the administration of the ITBS. The Switch List Test was given in May.

The Bryant consists of a list of fifty nonsense, common rule governed words which assesses growth and application in letter-sound correspondence knowledge. The Basal Word List consisted of over 500 core words, as determined by the respective publishers, from the first grade texts. At each time of testing the specific listed words used were determined by instructional progress through the series and the student's past test performance. The Switch List consisted of approximately 200 words which were unique to each series' core list, all words common to both Basal Lists having been eliminated. When this test was administered, children in basal series #1 (Economy) were asked to read list words from series #2 (Houghton Mifflin) and vice versa. It was assumed that while students may have seen these words in print, they had not received instruction based on them.

Analysis

Independent factors of interest (e.g., MRT scores, ethnicity, school, basal, classroom, etc.) were used to predict performance on the dependent measures (Iowa Test, Switch List, Basal Testing, etc.) through multiple regression analysis. The unique proportion of variance accounted for by each factor was determined by comparing full to reduced regression runs. In addition, means, SD's, Pearson correlations, and some one-way ANOVAs were obtained. All results discussed as significant (for both the ANOVAs and multiple regressions) had F values of at least .01.

RESULTS AND DISCUSSION

In determining what, if any, relationship existed among instructional materials, other factors, and the growth and application of letter-sound correspondence knowledge, several one-way ANOVAs were performed. Results indicated that mean correct response performance on the first Bryant test was significant by ethnicity, by school, and by basal. Anglo students, School #1 students, and students in the phonics basal performed best. Performance on Bryant #2 was not significant by either ethnicity or school, but the influence of the basal remained significant, with the phonics oriented basal students continuing their lead. However, none of these three factors was significant in predicting performance on Bryant #3 (see Figure 1).

Results of multiple regression analysis to determine the unique contribution of predictor variables on Bryant performance showed that the basal contributed 20% of the variance on Bryant #1 and 15% on Bryant #2 and #3. The Metropolitan showed a unique contribution of only 6% on Bryant #1, increased to 11% on Bryant #2, and jumped to 21% on Bryant #3. (We also looked at the

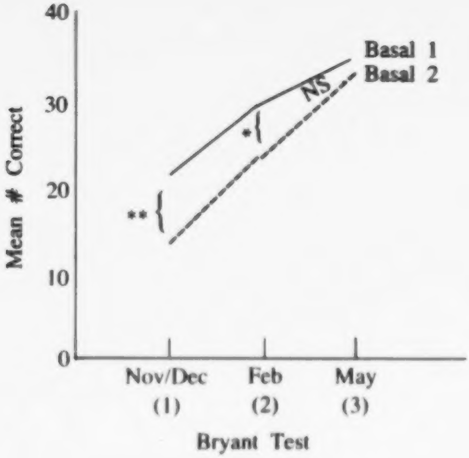


FIGURE 1. Influence of Basal Series on Bryant Performance.

** Bryant 1 by Basal F(1,86) = 13.8, $p<.001$;
* Bryant 2 by Basal F(1,88) = 5.5, $p<.02$;
NS Bryant 3 by Basal F(1,82) = .5.

contribution of each of the MRT subtests—auditory discrimination, visual discrimination, and language—and found none to be more predictive than the MRT total score for any of the Bryants.) The steady increase in prediction of the MRT and the decrease in the contribution of the basal suggests that as the year progressed, entry level readiness skills, as measured or reflected in MRT scores, became more important in the application of letter-sound correspondence knowledge. In addition, the drop in the unique contribution of the basal suggests that, while materials are very important at the beginning of reading instruction in first grade, they become less important as the year progresses. That is, the abstraction of letter-sound correspondence knowledge can initially be facilitated by materials which have lots of "regular" words, but sheer exposure to lots of words (as occurs in both series by the end of first grade) can also provide the "data" by which such knowledge develops. Such an interpretation is consistent with the results of the one-way ANOVAs which revealed an initial, but not continuing, advantage of the phonics series (see Figure 1).

Separate multiple regression analyses were performed to determine the unique contribution of factors predicting performance on the word list tests for each respective basal. For those children in the phonics oriented series, the MRT's independent power of prediction decreased throughout the year, showing an 8% contribution on Basal Test #1, 6% on Basal Test #2 and less than 1% on Basal Test #3. The unique contribution of class dropped from 19% in the Nov./Dec. testing to 15% on the May Basal Test #3. But the contribution of the Bryant rose steadily throughout the year from 10% on Basal Test #1, 14% on Basal Test #2 to 16% on Basal Test #3. These findings fit with the instructional focus of the phonics oriented series in which children were exposed to and received instruction in letter-sound correspondence. As a result, it appears that they learned the rules and could apply them to words from their series.

These same predictor variables showed a somewhat different pattern for performance in the more eclectic word pattern series. The MRT had less than 1% predictive value throughout the year. The independent contribution of class was a whopping 67% on Basal Test #1, dropped to 54% on Basal Test #2, and then fell to 13% on Basal Test #3. The unique contribution of the Bryant #1 on Basal Test #1 was less than 1%, Bryant #2 on Basal Test #2 was only slightly above 1%, while the independent contribution of Bryant #3 toward performance on Basal Test #3 jumped to 16%. These proportions suggest that once beginning readers are exposed to enough words to reach the saturation point for remembering words by sight (as likely occurred by Basal Test #3), letter-sound correspondence knowledge develops implicitly through exposure to print and becomes an important word recognition strategy. The strong initial impact of classroom further suggests that success in early "sight" learning is very dependent on what goes on instructionally within the reading group. In the phonics series, however, the materials themselves appeared to overshadow the impact of specific group instruction or practices.

In determining how important letter-sound correspondence knowledge is in the acquisition and performance of reading skills, the Basal Test #3, the Iowa Reading Total Score (reading comprehension and reading vocabulary), and the Switch List were used as the criterion variables. Once again the MRT total score, classroom, and performance on the Bryant #3 were used as predictor variables. As noted, in both basal series the independent contribution of the Bryant #3 toward performance on the Basal Test #3 was 16%. (The correlation of the Bryant #3 Test—regardless of basal series—with the performance on Basal Test #3 was +.67.) The unique contribution of class on Basal Test #3 was slightly less at 15% and 13%, respectively, in the two series.

Performance on the Switch List correlated +.43 with the Iowa Total score and the Bryant #3 performance correlated +.55 with the Iowa. The unique contribution of class toward predicting the Iowa total was 17%, but the independent contribution of the Bryant #3 was only 8%.

With regard to performance on the Switch List, class made a unique contribution of only 8%, but the Bryant #3's independent contribution to the total variance within Switch List performance was 22%. In addition, the Bryant #3 correlated +.72 with the Switch List.

These results suggest that the ITBS may be measuring something different from that which the Basal and/or Switch List are measuring. But it is clear that children who have acquired letter-sound correspondence knowledge do better both on reading the words that have been in their basals (Basal Test #3) *and*, perhaps more importantly, they do better reading words which they likely have not received instruction on or even read before (Switch List words).

CONCLUSIONS

Results of the current study thus suggest that, while the materials may be influential in facilitating early letter-sound correspondence knowledge, sheer exposure to lots of words also allows children to induce this information. Results further suggest that first graders who do indeed acquire such knowledge do better both in reading the words in their basals *and* reading words on which they have not received instruction.

REFERENCES

Bryant, N. D. *Diagnostic Test of Basic Decoding Skills*. Teachers College, Columbia University, 1975.

Durr, W., LePere, J., & Alsin, M. L. *The Houghton Mifflin readers*, 1st grade level. *Tigers, Lions, Dinosaurs, Rainbows, Signposts*. Boston: Houghton Mifflin Co., 1971.

Harris, T., Creekmore, M., & Greenman, M. *Keys to reading*, 1st grade level. *Pug, Sun up, Zip pop go, Green feet, Blue dilly dilly*. Oklahoma City: The Economy Company, 1972.

Hieronymus, A. N., Lindquist, E. F., & Hoover, H. D. *Iowa Test of Basic Skills*, Primary Battery, Form 7. Houghton Mifflin Co., 1980.

Nurss, J., & McGauvran, M. *Metropolitan Readiness Tests*, Level II. Harcourt Brace Jovanovich, Inc., 1976.

INA KATZ
San Bernardino State College

HARRY SINGER
University of California, Riverside

THE SUBSTRATA-FACTOR THEORY OF READING: DIFFERENTIAL DEVELOPMENT OF SUBSYSTEMS UNDERLYING READING COMPREHENSION IN THE FIRST YEAR OF INSTRUCTION

The substrata factor theory of reading postulates that as individuals learn to read they develop subsystems consisting of graphophonemics, morphophonemics, semantics and syntax. They can mobilize these subsystems for attaining comprehension in reading according to interactions between the purposes and resources of the reader and the demands of the print-stimuli (Holmes, 1976; Singer, 1962; Holmes and Singer, 1966). If instruction has emphasized one of these subsystems, then that subsystem is likely to be better developed than the other subsystems and be a strength in the pattern of subsystems mobilized for attaining reading comprehension (Barr, 1974; Biemiller, 1970; Chall, 1967; Singer, 1965). More specifically, the hypothesis to be tested in this study is that the development of the subsystems that can be mobilized for attainment of reading comprehension can vary as a result of differential emphases in methods of instruction.

Review of Previous Research

The Cooperative First Grade Reading Studies (Bond & Dykstra, 1967; Dykstra, 1967) attempted to find out if there are superior methods for teaching beginning reading. Analysis of the data from 27 independent studies indicated that no single method produced superior achievement in comprehension at the end of first grade. In fact, the results suggested that there was more variability in reading achievement within methods than between methods. In short, the differences between the methods in attainment of comprehension were not statistically significant. However, Bond and Dykstra did note that methodological emphasis on word identification yielded superior development in this subskill at the end of grade one. But a systematic analysis of subsystem development underlying reading achievement was not investigated in these studies.

A longer range study, conducted over two years, compared four instructional methods (Ruddell, 1968). The results indicated that at the end of the first year programmed instruction supplemented with instruction on language structure as related to meaning was superior. However, at the end of the second year, the basal program, which initially stressed whole word recognition and meaning components yielded significantly higher reading achievement. Although Ruddell concluded that the scope of the curriculum influences reading achievement, he did not investigate the effect of various instructional methods on differential development of the subsystems underlying reading achievement.

Katz and Singer (1976) reanalyzed the data from four of the first grade studies. They used analysis of covariance to determine whether one instructional strategy emphasizing letter-sound correspondence (ITA) and another strategy stressing whole word or a meaning-emphasis (basal reader) would yield statistically significant differences. The results indicated that an instructional treatment with a heavy emphasis on symbol-sound correspondence created a corresponding strength in the graphophonemic system. Only two of the comparisons of whole word basal reader strategies on development of the semantic subsystem were significant. However, the Early-to-Read (ITA) series, which ostensibly was employed as a graphophonemic method, used a meaning approach similar to that of the basal method; therefore, the findings comparing these methods were confounded. Also, the data from the original studies did not include any measure of syntactic subsystem development. Therefore, the First Grade data were not completely adequate for testing the hypothesis posed for this study.

Some evidence can be adduced to show that beginning readers use different subsystems to attain comprehension. For example, Barr (1974–75) studied responses of first grade children to ascertain comprehension. Half of them had been instructed with a graphophonemic emphasis and the other half with a whole word method. Her evidence indicated that the students' strategies were consistent with the methods emphasized in their respective classes.

Although beginning readers may start off with different strategies, Guthrie (1973) found that the subskills underlying reading comprehension become more interrelated as students learn to read. These subskills attain convergence by grade 3 or 4. Those students who do not attain this convergence are the poor readers in the group. Hence, convergence of the subability related to reading achievement is associated with increased reading comprehension.

Other researchers have also investigated the effect that variation in instructional methods has upon comprehension. Chall (1967) analyzed the results of 25 studies that compared whole-word and graphophonemic instructional approaches. Her findings suggested that initial instruction in letter-sound correspondence resulted in lower comprehension than did a whole word approach in first grade. However, initial instruction in letter-sound correspondence also led to early superiority in word recognition and better comprehension by the end of second grade.

More recently, Pflaum, Walberg, Karegianes, and Rasher (1980) used meta-analysis to synthesize the literature on instructional methods from 1965 to 1978. Although they found methods emphasizing sound-symbol blending were superior, they stressed that, in general, there were no differences among methods of teaching reading. However, in both of these studies, differential subsystem development as related to methods of instruction was not analyzed. Moreover, research design and data collection varied widely in the studies compared.

The research to date suggests that beginning readers may use different subsystems to attain comprehension (Adams, Anderson, and Durkin, 1978; Barr, 1974–75; Biemiller, 1970; Chall, 1967; Weber, 1970) and that instructional strategies influence the development of linguistic subsystems (Barr, 1974–75; Cohen, 1974–75; Katz & Singer, 1976; Katz, 1980). However, it does not clearly indicate whether variation in instructional methods does, in fact, differentially develop subsystems for attaining comprehension. Therefore, the purpose of this study was to test the instructional hypothesis that variation in instructional methods in the initial stages of formal reading development will differentially develop subsystems for attaining comprehension.

METHOD

Treatments

Four instructional treatments were designed to test the hypothesis. Each treatment differentially emphasized development of a particular linguistic subsystem. The four methods of instruction placed differential emphasis on development of the graphophonemic, morphophonemic, semantic, and syntactic subsystems.

Four curriculum guides, one for each instructional treatment, were prepared. Although the curriculum for each treatment was different, the instructional materials, tasks, and strategies used in each treatment were standardized so that all students were taught using, as much as possible, the same teaching strategies, materials, and group organization. Although these factors cannot be controlled completely we think that whatever variation may have occurred was not systematically related to any particular treatment. The basic vocabulary, which was the same for all treatment groups, was presented in chart story format.

Teachers

Four female teachers participated in a training period. They were instructed about (1) the concepts to be taught, (2) the materials to be used, and (3) the instructional strategies. During the treatment period, one of the experimenters met with the teachers weekly at their school sites and observed their instruction.

Each teacher taught two different instructional treatments on one school site. They chose the two treatments with which they felt most comfortable. Student assignment to teacher and treatment group was random. The order of treatment groups was counterbalanced.

Subjects

The subjects were 91 students from two elementary schools. They constituted the total academic kindergarten and first grade enrollment in these schools for the year 1979–1980. One child was subsequently eliminated from the study because of continuous absence due to a broken leg.

Both schools are located in middle to upper-middle class neighborhoods. However, as a result of voluntary desegregation and integration programs, the student population in these schools is heterogeneous. Analysis of the data indicated that the school itself was not a significant factor in relationship to student scores within treatment groups.

The treatment groups were formed within each school so that students could be assigned to each of the four treatment groups by means of stratified random sampling. Stratification variables were sex and reading ability.

Procedure

The students were already participating in reading instructional programs, some in *Open Court* (Hughes, Bermier, Gurren, Thomas, and Lebo, 1976) and others in *Ginn 720* (Clymer, Johnson, Jacobson, Shuy, and Torrance, 1975). Hence, all the students in this study received their usual reading instruction in their own classrooms plus the supplementary instruction specified in the instructional programs for this project. The supplementary reading instruction was divided into 17 sessions of 30 minutes per session.

The dependent variables for the first part of the study were linguistic subsystems and reading comprehension. Two instruments were used to measure the dependent variables: a standardized norm-referenced test battery, the Metropolitan Reading Instructional Tests (MRIT) (Farr, Prescott, Balow, and Hogan, 1978), and the criterion referenced tests (CRT) which were constructed for this study. Subtests of the MRIT used in this study include: (1) Reading Comprehension, (2) Auditory Discrimination, (3) Sight Vocabulary, (4) Phoneme-Grapheme: Consonants, (5) Vocabulary in Context, and (6) Word Part Clues. The CRT was constructed because a measuring instrument was needed that was more sensitive to specific achievement attained over a relatively short instructional period. The four subtests in the CRT are: graphophonemics, morphophonemics, semantics, and syntax. The items included in each subtest of the CRT assess only the knowledge of the content and the achievement of objectives of their respective instructional treatments. The corrected Spearman-Brown reliability estimates for these subtests range from .85 to .90.

RESULTS

A one-way analysis of variance was computed to measure the effects that the four subtests of instructional treatment had upon (1) development of each linguistic subsystem and (2) attainment of reading comprehension. Table 1 shows the means and standard deviations for the effects of treatments for each group on the CRT and on the MRIT. A comparison of the means in each row indicates that supplementary instruction had a differential but statistically significant effect in the expected direction on all four of the CRT subtests and on one of the MRIT subtests. Each group improved significantly over the other groups on its particular treatment.

These findings were confirmed by the ANOVA results, which revealed three significant findings:

1. Significant differences occurred between treatment groups on all four of the CRT ($p<.001$): Graphophonemics ($F(3,78)=6.37$), semantics ($F(3,78)=7.02$), morphophonemics ($F(3,78)=16.99$), and syntax ($F(3,78)=5.73$).
2. Only one significant difference between treatment groups ($p<.05$) was found for the subtests of the MRIT: Auditory discrimination ($F(3,78)=3.28$).
3. No significant difference was found between treatment groups on reading comprehension ($F(3,78)=3.28$).

Scheffé post hoc comparisons were used to locate the source of the significant differences between treatment groups on the CRT subtests and on the MRIT auditory discrimination subtest. The results of this analysis indicated:

1. On the graphophonemic subtest, comparison of the treatment groups showed that the graphophonemic group differed significantly from the morphophonemic group ($F(3,78)=5.72$, $p<.002$) and from the syntactic group ($F(3,78)=2.49$, $p<.07$).
2. On the semantic subtest, both the graphophonemic and the syntax groups differed significantly from the semantic ($F(3,78)=2.917$, 3.197, $p<.01$). The semantic treatment mean score was significantly higher than the mean scores of both the graphophonemic and the syntax groups. Although there was no significant difference between the semantic and the morphophonemic treatment

TABLE 1
Observed Cell Means and Standard Deviations For All Treatment Groups

			Treatment Groups			
Dependent Variable	**No. of Items**		**Grapho-phonemic**	**Morpho-phonemic**	**Semantic**	**Syntax**
Criterion Referenced Tests						
Graphophonemic	34					
Mean		24.01	27.91*	21.83	23.73	22.74
SD		5.66	4.30	6.46	4.54	5.34
Semantic	36					
Mean		25.43	21.82	26.26	29.50*	24.17
SD		6.71	6.34	6.33	3.73	7.66
Morphophonemic	28					
Mean		16.69	14.14	23.26*	14.14	15.00
SD		6.33	4.48	6.29	3.60	5.45
Syntax	20					
Mean		10.96	9.32	10.04	9.64	14.70*
SD		5.30	4.96	5.80	3.76	4.80
Metropolitan Reading Instructional Tests						
Comprehension	55					
Mean		20.54	19.23	22.96	18.41	21.44
SD		8.74	3.77	11.84	7.02	9.70
Auditory Discrim.	26					
Mean		18.90	18.41	18.75	17.49	21.20
SD		4.57	3.87	4.47	3.95	2.85
Sight Vocabulary	30					
Mean		20.62	20.75	20.88	19.36	22.31
SD		7.12	4.74	6.63	6.36	6.17
Phoneme/Grapheme	27					
Mean		19.67	20.26**	16.98	17.45	16.56
SD		5.90	4.06	5.96	6.06	4.95
Vocabulary in Context	24					
Mean		7.77	6.10	8.02	8.31**	8.02
SD		5.11	2.30	4.43	5.48	4.07
Word Parts	24					
Mean		8.85	7.92	10.03	7.11	9.39
SD		4.51	1.01	4.36	3.97	3.57

* Significantly different from the other treatment groups on the dependent variable for this row.

** Not significantly different from the other treatment groups on the dependent variable for this row, but the high Mean score for this treatment group is in the expected direction.

groups, the mean score of the semantic group was greater than the mean score of the morphophonemic group (29.50 and 26.26, respectively).

3. On the morphophonemic subtest, all of the other three treatment groups differed significantly from the morphophonemic treatment group ($p<.0001$). The mean score of the morphophonemic group was significantly higher than the mean scores of all other groups.

4. All the other treatment groups also differed significantly from the syntax group on the syntax subtest ($p<.01$). The mean score of the syntax treatment groups was larger than the mean scores of all other treatment groups.

The ANOVA also revealed a significant difference between treatment groups on the auditory discrimination subtest of the MRIT. A Scheffé comparison indicated that this auditory discrimination difference is the result of a difference ($p<.05$) between the semantic and syntax treatment groups. The mean score for the syntax group was 21.21; the mean score for the semantic group was 17.49.

DISCUSSION

The results of the ANOVA support the hypothesis that variation in instructional methods leads to differential development of the subsystems underlying achievement in reading comprehension. Furthermore, there were no differences on the particular subsystem among the three groups not receiving the particular supplementary instructional treatment.

Graphophonemics

Various theoretical views of reading acquisition place differential importance on instruction to develop linguistic subsystems. These theoretical views are based upon different assumptions. For example, the direct perception model assumes that beginning readers need to acquire only graphophonemic information to learn to read because they already have and are able to use the other subsystems, information necessary for processing written information. On the other hand, psycholinguistic models emphasize syntactic and semantic abilities; they assume that beginning readers need to utilize their linguistic capabilities for acquiring and using graphophonemic and morphophonemic knowledge in connected discourse. The beginning readers would have greater difficulty in acquiring this knowledge from isolated symbol-sound relationships.

Some empirical evidence has been adduced to support the claims made for each position (Chall, 1967; Goodman, 1965, 1972). For example, Chall (1967) found that first grade readers could acquire graphophonemic knowledge. Her findings have been used to support the position of the direct perception model. They can also be used to refute the assumption of psycholinguistic models that beginning readers are unable to acquire a graphophonemic subsystem when graphophonemics are presented in isolation.

The results of this experimental study are in agreement with the findings of Chall: beginning readers are, in fact, able to learn and use graphonemic rules for symbol-sound correspondences presented in isolation. The results of this study also indicate that beginning readers can learn morphophonemic rules under the same conditions.

Morphophonemics

Researchers have indicated that the morphophonemic subsystem facilitates comprehension of written material. Singer (1960, 1962) reported that morphophonemics is one of the predictors of comprehension at the fourth grade level. Rumelhart (1976) did not cite morphemics, as such, but he identified lexical level knowledge as one of the six knowledge sources used in interactive models. However, lexical level knowledge contains unbound morphemes, which we used in our morphophonemic treatment. But Rumelhart did not determine whether beginning readers can learn to identify morphemes. Singer (1965) did adduce evidence that readers at least at the third grade level use syllabication and at the fourth grade level use meanings of affixes. And Ruddell (1976) found that the addition of instruction in morphemes and syntax to graphophonemics (Sullivan's Programmed Instruction) improved first graders' comprehension on the Metropolitan Reading Test when compared with a control group, but Ruddell did not ascertain the specific contributions made by morphemes, syntax, and graphophonemics to reading comprehension. Our results, which showed that beginning readers can acquire morphophonemic ability, are also in agreement with the suggestion of Rumelhart that there is a morphophonemic knowledge source. Our results also extend Singer's findings to the beginning level reader.

Semantics

The results of previous research have indicated that readers as early as first grade can improve in semantic (Simons, 1975; Golinkoff & Rosinski, 1976; Clay, 1968; Weber, 1970) and syntactic knowledge (Weber, 1970; Levin & Kaplin, 1970; Goodman, 1972). Our results concur with these findings. They also indicate that even at the kindergarten level students can make progress in learning these subsystems.

SUMMARY

The substrata factor theory of reading postulates that readers mobilize their knowledge, abilities, and cognitive processes into working subsystems for solving problems in reading according to the interaction between their purposes and the demands of the printed material. These subsystems can be the six knowledge sources which the interactive model of reading (Rumelhart, 1976) posits can operate simultaneously, in parallel and/or in interaction with each other and with text-based data. The instructional treatments used in this study incorporated each of these knowledge sources. That is, feature knowledge, letter level knowledge, and letter cluster knowledge were subsumed in the graphophonemic treatment. Lexical level knowledge was assessed by the morphonemic subtest. Syntax and semantics measured other knowledge sources. The results of the ANOVA indicated that with differential supplementary instruction, beginning readers increased their development in each of the four linguistic subsystems: graphophonemic, morphophonemic, semantics, and syntax. In other words, a treatment with a heavy emphasis on a particular subsystem created a strength in the beginning reader's development of that subsystem. That is, differential methods develop quantitatively different subsystems for attaining reading comprehension. Hence, the data confirmed the hypothesis posed for this study: methods of instruction are functionally related to the pattern of subsystems that the beginning reader develops and can mobilize for attaining reading comprehension.

REFERENCES

Adams, J. J., Anderson, R. C., & Durkin, D. Beginning reading: Theory and practice. *Language Arts*, 1978, *55*(1), 19–25.

Barr, R. The effect of instruction on pupil reading strategies. *Reading Research Quarterly*, 1974–75, *10*(4), 555–582.

Biemiller, A. The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly*, 1970, *1*(1), 75–96.

Bond, G. L., & Dykstra, R. The cooperative research program in first grade reading instruction. *Reading Research Quarterly*, 1967, *2*(4), 5–142.

Chall, J. *Learning to read: The great debate*. New York: McGraw Hill, 1967.

Clay, M. M. A syntactic analysis of reading errors. *Journal of Verbal Learning and Verbal Behavior*, 1968, *7*, 434–438.

Clymer, T., Johnson, K., Blanton, W., Jacobson, M., Shuy, R., & Torrance, E. P. *Reading 720*, Massachusetts: Ginn and Co., 1975.

Cohen, A. S. Oral reading errors of first grade children taught by a code emphasis approach. *Reading Research Quarterly*, 1974–75, *10*(4), 616–650.

Dykstra, R. *Continuation of the coordinating center for first grade reading instruction programs*. Project No. 6–165, U. S. Department of HEW, September, 1967.

Farr, R. C., Prescott, G. A., Balow, I. H., & Hogan, T. P. *Metropolitan Reading Instructional Tests*, Primary 1, Form IJ. New York: The Psychological Corporation, 1978.

Golinkoff, R. M., & Rosinski, R. R. Decoding, semantic processing, and reading comprehension skill. *Child Development*, 1976, *47*, 252–258.

Goodman, K. S. A linguistic study of cues and miscues in reading. *Elementary English*, 1965, *43*, 639–643.

Goodman, K. S. The reading process: Theory and practice. In R. E. Hodges and E. H. Rudorf (Eds.), *Language and learning to read*. Boston: Houghton Mifflin, 1972.

Guthrie, J. Models of reading and reading disability. *Journal of Educational Psychology*, 1973, *65*, 9–18.

Holmes, J. A. Basic assumptions underlying the substrata factor theory. In H. Singer and R. Ruddell (Eds.), *Theoretical models and processes of reading*. Delaware: International Reading Association, 1976.

Holmes, J. A. & Singer, H. *Speed and power of reading in high school*. U. S. Dept. of Health, Education, & Welfare, Office of Education. (Cooperative Research Monograph No. 14). Washington, D.C., U. S. Government Printing Office, Superintendent of Documents, Catalogue No. F.S. 5.230:30016, 1966.

Hughes, A., Bernier, S. A., Gurren, L., Thomas, N., & Lebo, J. D. *Open court correlated language arts program*. La Salle, Ill.: Open Court Publishing Co., 1976.

Katz, I. C. *The effects of instructional methods on reading acquisition systems*. Unpublished dissertation, Riverside, California, 1980.

Katz, I. C., & Singer, H. *Effects of instructional methods on reading acquisition systems: A reanalysis of the first grade studies data*. Paper presented at the International Reading Association Conference, Anaheim, Ca., May, 1976, in ERIC. ED 122 227.

Levin, H., & Kaplan, E. Grammatical structure and reading. In H. Levin and J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Pflaum, S. W., Walberg, H. J., Karegianes, M. L., & Rasher, S. P. Reading instruction: A qualitative analysis. *Educational Researcher*, 1980, *1*(7), 12–18.

Ruddell, R. B. *A longitudinal study of four programs of reading instruction varying in emphasis on regularity of grapheme-phoneme correspondences and language structure on reading achievement in grades two and three*. Project No. 3099 and No. 78085, University of California, Berkeley, 1968.

Ruddell, R. B. Psycholinguistic implications for a systems of communication model. In H. Singer and R. Ruddell (Eds.), *Theoretical models and processes of reading*. Delaware: International Reading Association, 1976.

Rumelhart, D. E. *Toward in interactive model of reading* (Technical Report No. 56). San Diego: Center for Human Information Processing, University of California, 1976.

Simons, H. D. Transformational phonology and reading acquisition. *Journal of Reading Behavior*, 975, *7*(1), 49–59.

Singer, H. *Substrata-factor reorganization accompanying development in speed and power of reading at the elementary school level*. (U. S. Office of Education, Cooperative Research Project No. 2011.) University of California, Riverside, California, 1965.

Singer, H. Substrata-factor theory of reading: Theoretical design for teaching reading. In J. A. Figurel (Ed.), *Challenge and experiment in reading*. New York: Scholastic Magazines, 1962.

Singer, H. *Conceptual ability in the substrata-factor theory of reading*. Unpublished doctoral dissertation, University of California, Berkeley, 1960.

Weber, R. H. First graders' use of grammatical context in reading. In H. Levin and J. P. Williams (Eds.), *Basic studies in reading*. New York: Basic Books, 1970.

ROBERT E. LEIBERT
University of Missouri-Kansas City

A STUDY OF WORD ERRORS PRODUCED BY SECOND, THIRD AND FOURTH GRADE PUPILS READING THE DOLCH WORD LIST

Recently, a study of the effectiveness of playground graphics, as an incidental form of learning, was completed for a local school district using the Dolch words as a measure of the success of the project. The "Play and Learn Project" also provided an opportunity to analyze pupil responses to the list words.

Lists, such as the Dolch Word List (Zintz, 1981) are employed in classrooms to assure that basic word recognition will be developed. While the Dolch words are estimated to comprise up to 50 percent of the words in printed discourse, and researchers such as Harris (1980) have provided estimates of the grade level where these words are typically introduced in a basal series, there is no information about the difficulty of these words for pupils. Supported with data on the relative difficulty of the Dolch words a teacher could plan appropriate instructional and practice activities.

Word lists are also used as tests to estimate the strength of the word power of a pupil, and to assess the word recognition strategies being used by a pupil. The purpose of the current study was to determine the relative difficulty of the Dolch words for the project population, and to identify some of the general characteristics of incorrect word responses.

One hundred pupils were randomly selected from grades two and three and 94 pupils from grade four. These 294 children attended five schools located on the perimeter of the core city. The instructional program was the 1974 edition of the Houghton-Mifflin basal series.

Dolch words were typed in 24 point primer type and were divided into eleven lists of 20 words each. The lists were administered by college students, reading teachers and the investigator. After each list was completed, the pupil was asked to pronounce any word missed or skipped. When the testing was completed incorrect responses were analyzed in three ways: (1) as a distribution of total errors for each word, (2) by general characteristics and (3) by one distinctive feature—response variability.

A distribution of the Dolch words by the frequency with which they were missed indicates that the words *a, in, to* were not missed by any of the pupils in this population. Ranked by difficulty, the 110th word (*three*) was missed by 30 pupils and the mean number of errors for the 294 pupils was 25. Middle difficulty words were missed by about 10 percent of the population. At the other extreme, *far, gave, were* represented the most frequently missed words. *Were*, for example, was missed by 176 pupils of which 68 percent were grade two pupils, 63 percent were grade three pupils and 45 percent grade four pupils.

Because there is considerable spread in the difficulty of words, it is possible that pupils respond to easier words differently than more difficult words.

Table 1 shows a comparison of 30 average difficulty words (23 to 32 errors each) with 30 hard words (56 to 79 errors each). The following categories were constructed:

Refusals—words skipped or not attempted on either trial.
Incorrect Responses—an incorrect pronunciation attempt.

TABLE 1
Error Characteristics Between Average and Difficult Words for Second, Third and Fourth Grade Pupils

	2		3		4	
			Word Difficulty			
	Average	Hard	Average	Hard	Average	Hard
Total Errors	17*	35	06	22	05	12
Refusal	51**	47	20	38	30	29
Incorrect Responses	49	53	80	62	70	71
Words Corrected	35	28	39	28	38	46
Different Responses	47	42	51	38	60	46

*Percent of total possible responses
**Percent of total errors

Total Errors—a combination of refusals and incorrect responses.
Words Corrected—second trial corrections.
Different Responses—dissimilar incorrect responses.

As would be anticipated, the total number of errors decreased as grade level increased. The percentage of incorrect responses increased for each higher grade level reflecting the fact that more words were attempted by third and fourth grade pupils.

Somewhat less anticipated was the finding that for average difficulty words, pupils corrected approximately the same percentage of initial errors at each grade level. This finding also held true for harder words for second and third grade pupils.

Another unexpected finding was that the percentage of dissimilar responses was higher for grade four than for either grades two or three. Response variability is a curious phenomenon. Some words result in errors each one of which is different from the other, while other words produce few if any variations in incorrect responses.

Each of the 33 words listed in Table 2 produced an incorrect response which accounted for at least half of all incorrect responses for that word. Several observations can be made about the relationship of the response to the stimulus word. About half of these incorrect responses represent a shift in vowel sound such as: *wash* for *wish* and *fell* for *fall*. The remaining responses were graphically similar to the stimulus word as in *place* for *please* and *cold* for *could*. The probable influence of certain letters in the stimulus word over the response is most clearly portrayed by the word *there*. *There* (39 errors) produced the response *three* (14 times), and *three* (30 errors), also a Dolch word, produced *there* (10 times). But *their* (30 errors), another Dolch word, did not produce *three* as a response. Finally, half of these errors produced reciprocal errors. That is, an error response such as *no* (*no* for *on*) was also on the Dolch list and the list word *no* produced the original word *on* (e.g., ate-eat-ate and of-off-of).

Several other observations can be made regarding incorrect responses. Some errors were fairly obvious phonic attempts (e.g., *dōn* for *done*). A few responses, particularly those of beginning readers, could be classified as idiosyncratic or associational conflict responses (e.g., *of* for *that; play* for *ride*).

TABLE 2
Dolch Words Which Produced the Same Response for At Least Fifty Percent of All Incorrect Responses

Dolch Word/Number of Incorrect Responses/
Incorrect Response/Number of Responses/
Percent of Incorrect Response

no	(28)	on	(27)	96
good	(31)	go	(29)	94
came	(55)	come	(49)	89
gave	(107)	give	(94)	88
ran	(55)	run	(48)	87
use	(60)	us	(52)	87
know	(79)	now	(69)	87
thank	(79)	think	(69)	87
of	(55)	off	(45)	82
far	(71)	fair/fare	(58)	82
were	(151)	where	(124)	82
fast	(14)	first	(11)	78
am	(26)	I'm	(20)	77
cold	(46)	could	(35)	76
much	(19)	must	(14)	74
these	(52)	those	(38)	73
find	(41)	found	(29)	71
well	(52)	will	(37)	71
off	(26)	of	(18)	69
here	(45)	her	(31)	69
got	(19)	get	(13)	68
always	(33)	away	(21)	64
round	(41)	around	(26)	63
could	(32)	cold	(20)	62
very	(18)	every	(11)	61
please	(18)	place	(11)	60
wish	(50)	wash	(30)	60
found	(41)	find	(23)	56
why	(50)	way	(27)	54
myself	(13)	himself	(7)	54
wash	(43)	wish	(23)	53
must	(23)	most	(12)	52
has	(50)	his	(26)	52

Word lists provide information about the unaided recall of words. A second inspection of words is sometimes rejected in word list testing (Zintz, 1981), however, since the percentage of words corrected for this population ranged from 28 to 46, the omission of a second trial may result in counting words incorrect which are really known. The error distribution provides the teacher with information which should suggest words which are likely to be learned rapidly and words which may be expected to be more difficult to master.

The variety of responses for a word underscores how difficult it is to infer the recognition process from an incorrect response. Analyses by letter-sound characteristics assume that the pupil was applying a letter-sound strategy. Weber (1970) found that many of the incorrect responses of the pupils in her study were known words to the pupil. Ehri (1978) suggests that in the initial stages of mastering a word, identification may not be made on the basis of all the letters of that word. Noting the high graphic similarity between the stimulus word and many responses, and pupil knowledge of responses (reciprocal errors), another possible interpretation is that the incorrect response often represents the closest match that the reader can make with available stored words. It is possible that incorrect responses tell us more about what a pupil knows about a word than it tells us about what a pupil does not know.

REFERENCES

Ehri, L. C. Beginning reading from a psycholinguistic perspective: amalgamation of word identities. In Ehri, L. C., Baron, C. and Feidman, J. M. *The Recognition of*

Words, Newark, Delaware: International Reading Association, 1978, 13–19.

Harris, A. J., and Sipay, E. R. *How to Increase Reading Ability*, New York: Longman Inc., 1980, 372–374.

Weber, R. A linguistic analysis of first grade reading errors, *Reading Research Quarterly*, 1970, 3, 427–451.

Zintz, M. V. *Corrective Reading*, Dubuque, Iowa: Wm. C. Brown Company Publishers, 1981, 68–73.

BEVERLY J. OTTO
ELIZABETH SULZBY
Northwestern University

JUDGING THE EMERGENT READING ABILITIES OF KINDERGARTEN CHILDREN

The purpose of this paper is to describe the development and validation of an assessment tool for judging the emergent reading abilities of young children. The term "emergent reading ability," as used by Holdaway (1979; see also Sulzby, 1981a, 1981b), emphasizes the importance of early reading-like behaviors as being true literacy skills, not merely indicators of reading readiness in the traditional sense of predicting achievement in a yet-to-be-acquired skill.

Sulzby (1982) reported the development of a scale for use in assessing how close a child is to reading independently based upon the nature of the child's attempts to read from dictated and handwritten stories.[1] The scale is called Emergent Reading Ability Judgments (ERAJ). In dictation, a scribe writes the words which a child says in response to a "dictate" request. In handwritten composition, the child uses whatever writing system s/he prefers. After each type of composition, the child re-reads. The scale was developed from both the nature of the composition and the re-reading attempt.

Sulzby applied the scale to pairs of dictated and handwritten story-reading attempts of 24 kindergarten children; the study included a complete replication of story sets, leading to 96 judgments, 48 each for dictated and handwritten stories. Agreement between judges was high (96%) and the rank order correlation between the two replications each containing 48 attempts was rho $= .77$ (corrected for ties), $p<.01$. She reported a rank order correlation between average scores for four applications of the scale and the *Metropolitan Readiness Tests, Form P, Level 1*, of rho $= .37$ (corrected for ties), $p<.05$. She interpreted these scores to indicate that the scale did have acceptable reliability in scoring, consistency across time with the same children, and some relationship with scores derived from traditional measures for predicting reading achievement (readiness tests).

The Sulzby (1982) study was limited in scope to the consistency of the scale and to the relationship between the scale and a traditional readiness test. There was no assessment of the scale as a valid measure of emergent reading abilities of kinds other than reading dictated and handwritten compositions. Additionally, the scale was applied to the data from which it was derived so questions remain about whether there might be significant categories of behavior that were either not exhibited or not detected in the original sample.

These limitations and the need to explore further the emergent reading abilities of kindergarten children led to the current study. It was designed to address more specific issues of the reliability and validity of the Emergent Reading Ability Judgments for Dictated and Handwritten Stories (ERAJDHS; re-titled to distinguish it from other assessment tools being developed in BRDKAWL) (Sulzby, 1981a). The current study applies the scale to stories from a new sample of children whose overall emergent reading abilities were ascertained by other techniques.

This study addresses the question of whether the ERAJDHS scale covers the reading attempts of these children. If it covers the behaviors (a validity issue), can trained judges apply the scale with sufficient agreement and how reliable are repeated administrations of the scale of reading attempts of the same children over trials? How well does it discriminate subject groups chosen on overall emergent reading ability? Finally, how does it compare with a test of reading readiness?

METHOD

Subjects

Nine children were selected for case studies from among 24 children in a kindergarten classroom in an upper middleclass suburb north of Chicago, Illinois. The children were identified as being either high, moderate, or low in overall emergent reading ability at the beginning of the kindergarten year of 1980–81. This identification was based upon results of an interview study, "General Knowledge About Written Language," with teacher judgments used to refine the researchers' selections. Full procedures may be found in Sulzby (1981a).

The case study sample consisted of five girls and four boys, ranging in age from 4–11 to 5–11 in October. Three children fell into each group, with sex distribution not equal within groups. The high group consisted of three girls; middle group, of two girls and one boy; and the low group, of two boys and one girl. This distribution was reflective of the distribution of overall emergent reading ability within the 24-child classroom.

Data

The children's own compositions comprised the materials used for reading attempts. Dictated and handwritten story pairs were collected during the course of two different studies. Study II, "Real and Make-Believe Topic," contributed two complete sets, stories about learning to ride a big wheel. Study IV, "New Event Stimulus," yielded one set of stories about a race with novelty wind-up toys that triads of children participated in just prior to composition and re-reading. For purposes of this paper, those three story sets will be referred to as Trial 1, Trial 2, and Trial 3. Trial 1 took place in January and early February; Trial 2 took place in February; and Trial 3, in April and May.

Scoring

Two trained examiners scored the full collection of story/reading attempts. Initial scoring was done independently, using the Sulzby (1982) scale with conferences over disagreements. There were 54 story/reading pairs, with 27 dictated and 27 handwritten attempts. Judges were able to come to complete agreement but it should be noted that the agreement was conferred rather than independent for reasons that follow.

Initial scoring yielded 12 cases that did not fit the scale. These cases were of three kinds: one time the child produced what may be called a "pattern" story rather than the story requested; two times an examiner had given an illegitimate prompt during the re-reading attempts; and nine times the child's behavior did not fit the scale as originally worded. This final group consisted of one kind of behavior that was judged to be the equivalent of Sulzby's "five" category. The initial "five" category was defined as an instance in which the child's eyes were not on print but his/her story rendition was stable; the instances in which the scale did not fit were all times when the child's eyes were on print (not tracking) but his/her story was not stable. These behaviors were clearly more advanced than "four," and not so advanced as "six" in which the child is able to maintain eyes on print and a stable story, without actually reading the print independently. Table 1 gives the scale with this revision and with wording modifications for clarity.

After the scale was revised the scorers were easily able to classify 89% of the attempts, with the exception of the other three mentioned above. These three

TABLE 1
Emergent Reading Ability Judgments for Dictated and Handwritten Stories, Revised

Score Assigned	Behaviors Observed
(For 1–2, any production cannot be called a story or composition either by intention or by characteristics of the product.)	
1	No dictated nor handwritten stories; hence, no attempt to re-read. Child refuses to pretend-write.
2	In response to request to handwrite a story, no "story" is produced but there is some primitive evidence of reading and writing. In response to request to dictate a story, child says something that does not seem to be treated as a composed unit; it may consist of conversational characteristics either in wording or in turn-taking. It is either not re-read or rereading attempt is very primitive.
(For 3–7, stories or their equivalent are produced.)	
3	While stories are produced, child refuses to re-read. For own writing, child may say that the story "doesn't say anything," or, for dictation, "I can't read."
4	Child attempts to re-read but eyes are not on print. Story recited is similar to original but not stable.*
5	This seems to be a period during which the child cannot simultaneously maintain eyes on print and recite stable* story. Score "5" either if (a) child's eyes are on print but story recited is similar but not stable* or (b) child's eyes are not on print but story recited is stable.*
6	Eyes are on print but child is not tracking print. Story recited is stable.* For handwritten story, print may still be pretend-cursive or other writing system *if* the story accompanied the composition, either as shown by voicing during composition or by other evidence such as mixture of invention or reference to units by intention.
7	Child's eyes are tracking print, matching voice to print, "actually reading," independently, with attention to meaning. To be rated as a "7" story must consist of more than one complete clause; if less, then rank "6". Print can have mature forms of invented spelling or of combinations of other writing units as long as child treats individual units as stable.*

*Stable in reference to re-reading stories means that no clause-level units have been added, omitted, or placed out of composed sequence.

cases were discussed until the examiners agreed upon a placement for the attempts based upon inferences from surrounding evidence such as storytelling attempts, samples of aspects of reading, and other comments from the full protocols. One of these judgments resulted in a data point that seemed to be outside the general developmental pattern but the placement seemed to be consistent with the child's performance within the bounds of the scale; this was the "pattern" story in which the child dictated series of "I like my mom; I like my dad; I like my cat," sentences rather than the big wheel story.

Analysis

To examine the ability of the scale to distinguish between groups of children judged to be of differing ability, a 3 (Groups: high, moderate, low) × 2 (Mode: dictated, handwritten) × 3 (Trials: 1, 2, 3) mixed analysis of variance was conducted with modes and trials as repeated factors. Rank order correlations were calculated for each dictated/handwritten trial and for comparing the set of judgments with results of the *Metropolitan Readiness Test, Level I, Form P*, administered in November.

RESULTS

Results of the analysis of variance are shown in Table 2. These findings indicate that the scale differentiated between groups significantly, $F(2,6)=9.03$, $p<.025$. The means for the groups were as follows: High, 6.2; Middle, 4.4; and Low, 3.3. Two of the three children in the high group were at ceiling levels (total scores for six observations were 42 and 41 or 42), thus the differences between groups can be considered to be conservative.

In addition to group differences, the only other significant result of the

TABLE 2
Emergent Reading Ability Judgments for Dictated and Handwritten Stories (ERAJDHS) of Children of Differing Abilities Across Composition Mode and Trials

Source	SS	df	ms	F	p
Total	151.31	53	—	—	
Between	102.48	8	—	—	
Groups (G)	76.92	2	38.46	9.03	<.025
Error$_b$	25.56	6	4.26	—	
Within	51.83	45	—	—	
Mode (M)	8.16	1	8.16	4.61	NS
Trials (T)	6.03	2	3.01	1.45	NS
G × M	2.12	2	1.06	<1	NS
G × T	4.19	2	2.10	1.01	NS
M × T	6.79	4	1.70	5.31	<.01
G × M × T	2.77	4	.69	2.16	NS
Error$_1$	3.55	2	1.77	—	
Error$_2$	8.11	4	2.08	—	
Error$_3$	10.11	32	8	—	

analysis of variance was the interaction between Mode and Trials, $F(4,32)=5.31$, $p<.01$. Figure 1 shows that interaction graphed; dictated story reading attempts tended to be scored slightly higher than handwritten stories for the final two trials, but for Trial 1 they were quite a bit higher (mean of 5.2 compared with 3.4).

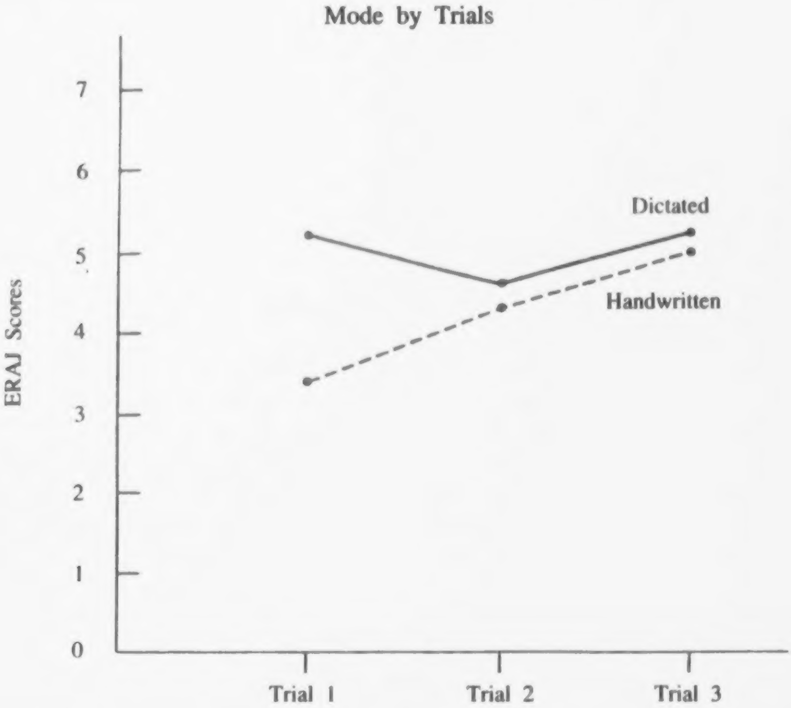


FIGURE 1. Mean Emergent Reading Ability Judgments for Dictated and Handwritten Stories.

Another question about group differences can be addressed by examination of the means for the nonsignificant Groups by Trials interaction. The question is two-fold: (1) were groups distinguished across time, and (2) did children increase in Emergent Reading Abilities for Dictated and Handwritten Stories over time? While the main effect for groups answered the first part of the question in the affirmative, the means displayed in Table 3 show that, while the high group was almost at ceiling and stayed stable over time, the low and middle group mean scores increased across trials.

The rank order correlations (corrected for ties) between the pairs of dictated and handwritten stories were as follows: Trial 1, rho=.83 ($p<.01$); Trial 2, rho=.58 (NS); and Trial 3, rho=.73 ($p<.05$). The correlation between the total scores for the ERAJDHS for the six sessions and the total raw scores on the *Metropolitan Readiness Test* was rho=.80 ($p<.05$, uncorrected since there was only one tie).

TABLE 3
Mean Emergent Reading Ability Judgments for Dictated and Handwritten Stories: Groups by Trials

Groups	Trial 1	Trial 2	Trial 3
High	6.3	6.2	6.2
Middle	3.8	4.2	5.2
Low	2.8	3.2	4.0

These correlations indicate that the rank order of subjects between the paired administrations of the scale tended to be maintained. Additionally, the rank associated with the subjects' total scores for the six sessions tended to be close to the rank of subjects' scores from the Metropolitan.

DISCUSSION

The scale, Emergent Reading Ability Judgments for Dictated and Handwritten Stories, was able to be applied to repeated reading attempts of kindergarten children. This application required one revision in order to capture behavior not covered by a previous version of the scale. Revision of the scale resulted in an assessment tool which could be used by trained examiners with high agreement.

The scale is limited to children's reading attempts using their own compositions, dictated and handwritten, as the texts. A further limitation of the scale is that people using it must be knowledgeable about the nature of young children's writing systems. The scale requires the assessment both of the composition and of the re-reading attempt. It is thus a one-to-one assessment tool. Finally, the scale only makes gross discriminations of ordinal level. It does not give finer diagnosis of the child's attempts as the child nears independent reading. Thus it may imply a more rigid sequence than a closer examination of exact reading-like behaviors reveal.

Nevertheless, the analysis of variance over three repeated administrations of the scale to dictated and handwritten story reading attempts indicated that the scale discriminated between groups of children who differ in overall emergent reading abilities. While the main effect for trials was not significant, there was also a trend in the data for children to improve across time, particularly for handwritten stories.

In addition to the ability of the scale to discriminate groups of children the scale also showed moderate and, in two of three instances, significant ability to rank subjects similarly across dictated and handwritten forms within a given trial. This correlation is worth noting, given the mode by trial interaction (see Figure 1). While the dictated mean for Trial 1 was significantly higher than the handwritten mean, the ranks of scores for those two administrations were significantly related.

Finally, even though the two measures are devised upon different bases for different purposes, the scale and a traditional readiness test also tended to rank children similarly. The correlation for this comparison was the children's total ERAJDHS scores across four to five months and the total raw score for the readiness test. It would be surprising and somewhat suspicious if the two measures were totally unrelated since both purport to measure some facets associated with written language.

It should be remembered that the correlations reported are for ranks and are for nine children. Hence, the significance level is more important than the magnitude. Sulzby (1982) reported significant rank order correlations for a different and larger sample (N = 24).

The Emergent Reading Ability Judgments for Dictated and Handwritten Stories has two potential uses. It can be used, along with other tools, in research in early literacy development. It can also be used by teachers of young children to provide one measure of ongoing emergent reading and writing abilities. In both kinds of use, it is limited to being a gross measure of structural characteristics. It needs further refinement and should be used in combination with other assessment tools; nevertheless, it seems to provide some assurance of being one reliable and valid assessment tool for reading behaviors of young children who are not yet reading independently.

NOTES

1. The term "story" is used here to indicate the composed text, regardless of any genre classification, because the term "story" was used in directions to the children. For a discussion of text types in relation to emergent reading abilities, see Sulzby, E. *Kindergarteners begin to read their own compositions: Beginning readers' developing knowledges about written language project*. Final report to the Research Foundation of the National Council of Teachers of English. Evanston, IL: Northwestern University, 1981, or the conference version of this paper available from the authors.

ACKNOWLEDGEMENTS

The research reported in this paper was supported in part by a grant from the National Institute of Education, NIE-G-80-0176, to the second author. The statements made are solely the responsibility of the authors. The authors thank the children and their teacher, who must remain anonymous. We also thank Susan Anderson, Beverly Cox, and Dr. Norman Bowers for their assistance.

REFERENCES

Holdaway, D. *The foundations of literacy*. Sydney: Ashton Scholastic, 1979.

Sulzby, E. *Beginning readers' developing knowledge about written language*. Interim report of the National Institute of Education. Evanston, IL: Northwestern University, 1981. (a)

Sulzby, E. *Kindergarteners begin to read their own compositions: Beginning readers' developing knowledges about written language project*. Final report to the Research Foundation of the National Council of Teachers of English. Evanston, IL: Northwestern University, 1981. (b)

Sulzby, E. Oral and written language mode adaptations in stories by kindergarten children. *Journal of Reading Behavior*, 1982, *14*, 51–59.

MICHAEL R. SAMPSON
East Texas State University

A COMPARISON OF THE COMPLEXITY OF CHILDREN'S DICTATION AND INSTRUCTIONAL READING MATERIALS

Children constantly seek to bring meaning and order to their world. They spend the first few years of their lives experimenting with and practicing language, building their store of word-meanings and concepts on personal experiences and interaction with others. Halliday (1975) describes this process as "learning how to mean." Because language is based on experience and meaning, children entering school usually possess well-developed vocabularies and an intrinsic understanding of how language functions.

In order for learning to continue at the same pace or to accelerate, this pattern of associating experiences and meaning with language must not be broken when children begin formal instruction in school. Unfortunately, this link between language, meaning, and reading has not been maintained in the past. Faced with the challenge of teaching children to read, teachers usually resort to the use of a basal reading program. The preprimer and primer stories of basal reading series are not known for their literary quality. In fact, according to research summarized by Giles (1966), children exposed to preprimers and primers actually experience vocabulary regression, rather than growth.

Some publishers of basal reading series are now contending that the days of simplistic, low-vocabulary beginning reading materials are over. However, it is of great importance to raise the following basic questions: (1) To what degree have beginning reading materials been improved, or, how far have we come since the "Dick and Jane" days? (2) Are children now being exposed to materials that are meaningful and reflective of a literary level that is commensurate with the students' own language abilites?

The purpose of this study was to determine the relationship between the language level of first-grade children and the language level of selected preprimers used for beginning reading instruction. Of particular interest was the difficulty level of the basal texts in terms of the philosophical orientation (phonics, skills, whole language; Rhodes 1979) of the series toward the reading process.

METHOD

Subjects

All 15 first-grade classrooms in a rural school district participated in this study; 50 students were randomly selected as subjects from these classrooms. The socio-economic status of the students ranged from low to upper middle class. The study was conducted in October of the 1981–1982 school year.

Materials

Seven preprimers from seven basal reading series were represented in this study. Of these, two were representative of a phonics approach to the teaching of reading, four were representative of a skills approach, and one was representative of a whole language approach. The second preprimer in each series was used in the study.

Instruments

The Allen Language Assessment Formula (Allen 1979) was used to assess the language complexity of both the preprimers and the subjects' dictated stories. Developed with careful consideration to basic psychological and linguistic principles, this formula recognizes that literary quality is raised when authors use adverbials, figuratives, and a variety of creative constructs in their writing.

Procedure

Subjects dictated stories individually to the researcher. The stories were based upon an experience or a fantasy provided by the students. Stories were written using the students' home-rooted language; no grammatical or structural corrections were made. The dictated stories were later analyzed using the Allen Language Assessment Formula.

Textbooks were also analyzed using the Allen Language Assessment Formula. The short length of preprimer texts allowed for a total analysis of each book.

RESULTS

No statistically significant difference existed between the students' language level ($\bar{X}=360.52$, $SD=59.32$) and textbook Ā ($\bar{X}=346.60$; $t(49)=1.66$, $p>.05$), textbook E ($X=354.83$; $t(49)=1.47$, $p>.05$), and textbook G ($X=365.17$; $t(49)=.555$ $p>.05$). However, the language level of the subjects was significantly superior to the language level of textbook C ($X=341.2$; $t(49)=2.30$, $p<.05$); textbook D ($X=321.2$; $t(49)=4.69$, $p<.001$); and textbook B ($X=321.4$; $t(49)=4.66$, $p<.001$). A statistically significant difference between the students' language level and textbook F favored the textbook ($X=377.85$; $t(49)=2.07$, $p<.05$).

DISCUSSION

The results of this study indicated that the relationship between the complexity level of children's dictation and the complexity level of preprimer texts varies from series to series. Unfortunately, in five of the seven preprimers surveyed, the mean scores of the students were higher than the mean scores of the texts, with the differences being statistically significant in three of the cases. This is, of course, an indication that students are not being fully challenged and provided opportunities for language growth when instructed with such preprimer materials. Instead, they are subjected to less sophisticated vocabulary and language patterns than they themselves use.

In terms of text complexity and the philosophical orientation of the series toward the reading process, one of the phonics model preprimers was simpler than students' vocabulary and language patterns (D), and one was more complex (F). Phonics series are well known for their instructional focus on letter-sound relationships. Consequently, they frequently contain stories with minimal content and unnatural, stilted sentences. Therefore, it was surprising to find a phonics textbook (F) with complex language structure and well-developed stories. A closer examination of the two phonics preprimers revealed that the stories in text D were written to feature certain letter-sound relationships, whereas stories in text F were not so designed. This was an encouraging finding, because it means that teachers who use series F are providing students with good models for language.

All four of the skills model textbooks (A,C,D,E) ranked lower in complexity than did the language of the students. The language sophistication of the whole language series (G) was more complex than the language sophistication of the students. In addition, the semantic and syntactic repetitions found in some of the stories made the text predictable, and thus easier to read.

The exposure of children to good literature is of great importance. Because basal readers are usually provided for all students, it is imperative that a wise, discerning decision be made when selecting a series. Materials and programs of instruction which limit beginning reading experiences to repetition of high-frequency words may deprive students of a valuable opportunity for vocabulary growth and actually promote vocabulary regression. Children need to be exposed to language that better utilizes nouns, adjectives, verbs, and adverbs, and stories that have such features as conversation, figuratives, adverbials, and other creative constructs.

In summary, many preprimers continue to present to children language which is inferior in vocabulary and style to the language used by children. Thus, it may be concluded that one of the best sources of materials for reading instruction resides in the language strengths of the students. Teachers who utilize this language strength through the dictation process and student-authored books are wisely capitalizing on these language strengths.

REFERENCES

Allen, R. V. *Allen Language Assessment Formula*. Unpublished manuscript, University of Arizona, 1981.

Giles, D. E. *The Effects of Two Approaches to Reading Instruction Upon the Oral Language Development of First Grade Pupils*. Unpublished doctoral dissertation, North Texas State University, 1966.

Halliday, M. A. Learning how to mean. In E. Lenneberg and E. Lenneberg (Eds.), *Foundations of Language Development: A Multidisciplinary Approach*, Vol. 1. Academic Press, 1975.

Rhodes, L. K. Comprehension and Predictability: An analysis of beginning reading materials, *New perspectives on comprehension*, 3, 1979, Indiana University.

PATRICIA A. ALEXANDER
Texas A & M University

WHEN ARE QUALITATIVE RESEARCH METHODS USEFUL IN READING-RELATED INQUIRY?

As indicated by its emphasis in the literature and in formal instruction, there is little doubt that the quantitative methodology dominates reading research (Rist, 1977; Weintraub & Farr, 1976). Yet, there are those in reading who have broken free of this "conceptual cul-de-sac" (Rist, 1980, p. 8) and have sought out alternative, qualitative models of research. Among the possible explanations for the emerging interest in the qualitative methodology is the apparent disappointment among some investigators with the little that has been accomplished in the areas of cognition and instruction. What seems to heighten their disappointment is the realization that the processes that concern them are so abundantly present in the academic setting (Seibert & Snow, 1965).

The concept of reading, itself, has undergone transformations that have had an impact on the research methodologies employed. These transformations may furnish an alternative explanation for the implementation of qualitative methodologies in some reading research. With the acceptance of the reading process as an interaction between reader and text it has become necessary to look to other disciplines such as cognitive psychology, sociolinguistics, and cultural anthropology to achieve greater understanding of this complex phenomenon. As reading becomes more influenced by the contributions of other disciplines, it also must become more affected by the various methodologies these disciplines employ to achieve their ends. Further, as the concept of reading expands into the domains of other disciplines new and varied questions relative to the reading process will be raised. It should be these questions that determine the method of research applied (Wardrop, 1976). After all, researchers should see themselves as solvers of puzzles, willing to try any number of approaches or strategies to achieve results, rather than as preservers of some paradigm (Kuhn, 1970). This paper poses the question of when and for what purposes qualitative research methods should be used to address reading-related questions. Qualitative research, as broadly defined here, encompasses a wide array of methodologies. These methodologies range from the strict adherence to traditional, ethnographic techniques, where intervention and implications are excluded, to the insertion of descriptive data within an otherwise standard, empirical study.

One primary purpose for the application of qualitative methods is to investigate those cognitive "black boxes", the unobservable processes that result in observable behaviors. The consideration of subjects' verbalization during task completion, whether central or supplementary to the study's main purpose, demonstrates the use of qualitative analysis for speculation as to *process*. As Snow (1974) noted, intra-experimental observation is an essential component in research and investigators must begin to take "detailed account of what goes on inside the Ss inside the experiment" (p. 279). The implementation of this purpose can be viewed in the recent work by Fillmore (1981) and Miller and Kintsch (1980). In the Miller and Kintsch investigation of readability, the researchers examined the match between reader and text by having subjects verbalize their expectancies as to evolving events at designated points in the reading. These verbalized predictions were then compared to the actual events in the text and the reading continued. While the researchers quantitatively analyzed reading time as an indicator of text readability, they qualitatively examined the verbal reports of subjects as evidence of the cognitive processing that was taking place.

In a similar fashion, Fillmore (1981) studied the relationship between demands of reading comprehension test items and reader's abilities. In this ideal/real reader research, Fillmore requires subjects to verbalize their understandings of item content after the addition of each subsequent phrase. From these verbalizations Fillmore has established levels of envisionmment—levels of knowledge and strategic processing at which comprehension success or failure is more likely to occur. Other research studies in reading-related areas have also employed intra-task verbalizations as a source of data (e.g., Garner & Alexander, in press; Olshavsky, 1976–77).

Another instance when the use of qualitative methods may be more applicable in reading research is when the natural interactions among variables within a "real world" context are of concern. Bronfenbrenner (1977) contends that no study can be viewed as context free. Even laboratory experiments have their own special effects and their own "ecological context". Rather than seeking to sterilize the situation, those undertaking more qualitative research can invite the natural inference that is part of the designated context (Guba, 1978), carefully describing that context through ethnographic techniques.

The sociolinguists have been very influential in underscoring the potent relationship between language function and social context as illustrated by the work of Cazden (1970), Erickson (1981), and Cherry (1980) among others. The perspective of the sociolinguists is argued well by Mishler (1978) who states that all human action is context-dependent, even that which relates to language functions and uses. However, the paradox, according to Mishler, is that this common sense principle has been ignored in the main tradition of social science research and theory. To achieve their purposes, therefore, sociolinguists have modified tradition, and have expanded their research arsenal to include naturalistic observational techniques within conventional and non-conventional ethnographic frameworks. One recent study that has combined the concern of sociolinguistics and reading was Wilkinson, Clevenger and Dollaghan's (1981) investigation of interactions among first graders in reading groups. In this study the communication patterns of students working independently on a specified instructional task were qualitatively examined. A secondary purpose of this research was to demonstrate the applicability of the sociolinguistic methodology to classroom communication—a purpose that the Wilkinson et al. (1981) research seems to have accomplished.

The implications of treatment-situation interactions have also been assessed in the research of Cronbach and Snow (1977). Aptitude-Treatment-Interaction, ATI, considers the relation between aptitude-achievement performances and how these can be moderated or modified by instructional treatment variables. It is Snow's contention that "we must stop acting as if cognitive abilities and achievements were arrays of separate mental garbage cans" (1980, p. 44), and begin to explore the dynamics of learning.

Qualitative methodologies may also be more suitable than quantitative models in the investigation of constructs which are complex and dynamic. If the behaviors to be studied are assumed to be multi-dimensional and probabilistic then trying to strip them down to the minimum to achieve control may be potentially self-defeating (Snow, 1980). A prime example of an area that is both complex and dynamic is the process called reading. Unless researchers perceive the reading process "as a whole dynamic area" they are "likely to study bits and pieces that may have little or no significance in furthering understanding of the reading field" (Robinson, 1976, p. 8).

Erickson (1977) has listed a number of social relations within the instructional environment that could be operationalized and monitored ethnographically. These relations included: the amounts of time students attend to the teacher; the topical relevance of discourse between student and teacher, and between student and student; teacher assessment of students' intellectual competence on the basis of social performance; and, the regularity of classroom activity rhythms. Certainly, the classroom setting contains a myriad of discriminable verbal and non-verbal behaviors, ignored by quantitative methods, but observed and analyzed by qualitative techniques (Dunkin & Biddle, 1974).

There are those situations when there is inadequate theory or research to indicate the parameters of the topic of interest or the parameters remain too vague to permit hypothesis formation. For such situations an exploratory qualitative approach may be preferable to a confirmatory quantitative approach. Conventional empiricists by the nature of their methodology must become reductionists, focusing on a controllable number of variables to the exclusion of others. By contrast, naturalistic inquirers assume a posture of ignorance about the phenomenon under study, allowing variables and research questions to emerge freely from the data.

Shuy and Griffin (1981) studied functional language in schools. The researchers employed a qualitative methodology because of their assumption that "language data, in this case school language, should be as naturalistic as possible" (p. 273). Studying language in its naturalistic setting, according to Shuy and Griffin (1981) is a necessary part of deciding how to plan an experiment or interpret its results. The examination of a phenomenon within its naturalistic setting also increases the likelihood that the experiment is "ecological valid"; that is, that the situation described is representative of the population of situations to which researchers wish to generalize (cf. Snow, 1977).

Durkin's (1978–79) research into comprehension instruction was another attempt to qualitatively unearth the occurrences in the classroom. What Durkin's observational investigation of reading classes revealed was that very little comprehension instruction actually took place in the elementary schools included in her study. This suggests a need to look closer at instructional priorities in reading classes. Additional investigations of comprehension instruction may find it beneficial to combine quantitative and qualitative methodologies in the research of this and other reading-related phenomena. As the ethnographic principle of triangulation suggests, if similar results can be reached from differing points of departure, as in the case of diverse research methodologies, the strength of the findings can only be enhanced.

The point to be made here is that this paper does not advocate the qualitative method to the exclusion of all others. There is no one best method of research. The suitability of research methodologies can only be determined by the question they attempt to address. What is advocated is that there are situations, such as those previously described, and others abundantly present in the instructional setting that can be better examined through qualitative means. As reading researchers we can only hinder our progress "by a continued fixation" on what is good or bad about one approach or another (Rist, 1977, p. 42). What is needed is a sense of detente among qualitative and quantitative researchers—a willingness to set aside methodological biases in order to achieve a common objective, greater understanding of the process called reading.

REFERENCES

Bronfenbrenner, U. Toward an experimental ecology of human development. *American Psychologist*, 1977, *32*, 513–531.

Cazden, C. B. The neglected situation in child language research and education. *Journal of Social Issues*, 1970, *26*, 35–60.

Cherry, L. J. A sociolinguistic approach to language development and its implications for education. In [illegible]. Garnica & M. King (Eds.), *Language, children, and society*. New York: Pergamon Press, 1980.

Cronbach, L. J., & Snow, R. E. *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington, 1977.

Dunkin, M., & Biddle, B. *The study of teaching*. New York: Holt, Rinehart, & Winston, 1974.

Durkin, D. What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 1978–79, *14*, 481–533.

Erickson, F. Some approaches to inquiry in school-community ethnography. *Anthropology and Education Quarterly*, 1977, *8*, 58–69.

Erickson, F. Timing and context in everyday discourse: Implications for the study of referential and social meaning. In W. P. Dickson (Ed.), *Children's oral communication skills*. New York: Academic Press, 1981.

Fillmore, C. J. *Ideal reader–Real reader*. A paper presented at the annual meeting of the Georgetown University Round Table, Washington, D.C., 1981.

Garner, R., & Alexander, P. Strategic processing of text: an investigation of the effect on adults' question-answering performance. *Journal of Educational Research*, in press.

Guba, E. G. *Toward a methodology of naturalistic inquiry in educational evaluation*. (CSE Monograph Series in Evaluation, No. 8), Center for the Study of Evaluation. Los Angeles: University of California, 1978.

Kuhn, T. S. *The structure of scientific revolutions* (2nd ed.). Chicago: The University of Chicago Press, 1970.

Miller, J. R., & Kintsch, W. *Recall and readability of short prose passages*. A paper presented at the annual meeting of the American Educational Research Association, Boston, April 1980.

Mishler, E. *Meaning in context: Is there any other kind?* A paper presented at the annual meeting of the American Educational Research Association, Toronto, April 1978.

Olshavsky, J. E. Reading as problem solving: An investigation of strategies. *Reading Research Quarterly*, 1976–77, *12*, 654–674.

Rist, R. C. Overview: On the relations among educational research paradigms: From disdain to detente. *Anthropology and Education Quarterly*, 1977, *8*, 42–49.

Rist, R. C. Blitzkrieg ethnography: On the transformation of a method into a movement. *Educational Researcher*, 1980, *9*, 8–10.

Robinson, H. M. Identifying and planning reading research. In R. Farr, S. Weintraub & B. Tone (Eds.), *Improving reading research*. Newark, Delaware: International Reading Association, 1976.

Seibert, W. F., & Snow, R. E. *OASIS: A Methodology for Instructional and Communications Research*. A paper presented at the annual meeting of the American Psychological Association. Washington, D.C., 1965.

Shuy, R. W., & Griffin, P. What do they do at school *any* day: Studying functional language. In P. W. Dickson (Ed.), *Children's oral communication skills*. New York: Academic Press, 1981.

Snow, R. E. Representative and quasi-representative designs for research on teaching. *Review of Educational Research*, 1974, *44*, 265–291.

Snow, R. E. Aptitude and achievement. *New Directions for Testing and Measurement*, 1980, *5*, 39–59.

Wardrop, J. L. Design problems in reading research. In R. Farr, S. Weintraub & B. Tone (Eds.), *Improving reading research*. Newark, Delaware: International Reading Association, 1976.

Weintraub, S., & Farr, R. Introduction. In R. Farr, S. Weintraub & B. Tone (Eds.), *Improving reading research*. Newark, Delaware: International Reading Association, 1976.

Wilkinson, L. C., Clevenger, M., & Dollaghan, D. Communication in small instructional groups: A sociolinguistic approach. In P. W. Dickson (Ed.), *Children's oral communication skills*. New York: Academic Press, 1981.

JOSEPHINE SPIVACK GOLDSMITH
Rutgers University

MARK J. NICOLICH
Rutgers Medical School

EDWARD J. HAUPT
Montclair State College

A SYSTEM FOR THE ANALYSIS OF WORD AND CONTEXT-BASED FACTORS IN READING

This paper presents a system for the coding and analysis of oral reading errors. Oral error analysis remains a highly useful method which potentially could measure the joint contributions of word and context based factors to given moments of reading. For this reason the technique is applicable to investigate many issues posed by interactive models of reading (Rumelhart, 1977; Stanovich, 1980).

Important basic work in developing the technique was done by K. and Y. Goodman and their colleagues (Goodman, K. S., 1968; Goodman, K. S. & Goodman, Y. M. 1977). These researchers applied the concept of the taxonomic model to the area of oral error analysis. They developed two scales: the Reading miscue inventory (RMI; Goodman & Burke, 1972) designed for classroom and clinical diagnosis, and the *Taxonomy* (Goodman, K. S., 1969; Goodman, K. S. and Burke, C. L., 1969) designed for research use. The RMI classifies the similarity of the word in the text to the error on nine scales measuring various aspects of word and context based reading. In contrast, the *Taxonomy* incorporates 28 scales designed to produce an in depth analysis of reading errors. The *Taxonomy* is an elaborate system with wide ranging categories including such elements as a system for tracing recurring errors; a scale to measure the language unit involved in an error, from the submorphemic level to the sentence; and a method to measure the influence of cues from peripheral vision.

The *Taxonomy* was designed to produce descriptive studies using small samples. Perhaps for that reason the system is unsuitable for the analysis of a large data base. Two limitations in this regard are as follows:

(1) Levels within scales differ qualitatively, but are not constructed at an ordinal level of measurement. Goodman's analyses are presented as percentages, rather than as measures of variance. Adaptations are needed to allow the application of standard techniques of data analysis.

(2) Passages differ from study to study. Therefore, generalization over studies becomes problematic.

Furthermore, Leu (1982) has presented three additional methodological difficulties which plague studies of oral error analysis. First, the choices and definitions of categories, as well as the number of levels within categories, differ widely from study to study. Second, the effect of passage difficulty on error type has been inadequately explored. Finally, no current system is capable of measuring the joint contributions of several sources to error variance.

In spite of these limitations, of systems currently available the technique offers major advantages. First, the method is more generalizable to a classroom reading situation than are many laboratory techniques since it involves reading in context with the use of typical school-like materials. Second, it is one of few techniques which has the potential to permit us to partial out the relative influences of word and context both for individual errors and total reading performance.

Given the limitations outlined above, modifications of the present methods were needed to allow researchers to apply the technique to a wide range of research questions.

DESIGN OF THE SYSTEM

The system presented here (Haupt & Goldsmith, 1980) is such an adaptation of Goodman's taxonomic concept, with modifications which incorporate related work of other researchers (Hood, 1976–77; Venezky, 1970).

The use of the categorical model and several of the scales draw from Goodman's concept of miscue analysis. The highly operational system developed by Hood for coding of contextual appropriateness has been incorporated with an adaptation to allow for separate ratings for syntactic and semantic appropriateness. Venezky's orthographic categories have been adapted and simplified to produce the grapheme categories for the scale of phonemic similarity.

In answer to Leu's first criticism directed at the definitions of scales and levels, we have attempted to provide categories which are defined unambiguously and can be applied consistently. Further, the system incorporates tests of consistency and of the underlying structure of the scales through the use of reliability studies and factor analysis.

Level of measurement. Our adapted scales measure the similarity of each error to the target word in the text at an ordinal level, with lower values indicating greater similarity. In general, the scales range from 1 to 4 with the exception of phonemic similarity, a proportional scale, which ranges from 1 to 9. The development of ordinal scales enables us to analyze the data using standard statistical techniques.

Passage development. The issues of passage difficulty, Leu's second criticism, and of passage generalizability are addressed by the development of standard passages. In this way we have controlled passage difficulty rather than testing its effects, an issue reserved for later investigation. (See Leslie [1980] for a direct investigation of this issue.) For each grade tested (two, four and six) both a narrative and a science based passage were developed. The passages are approximately 400 words in length and were selected from standard classroom materials to form coherent selections. In accordance with the low frustration rate prescribed for passages used for oral error analysis (Y. Goodman & Burke, 1972), we selected and pre-tested passages to achieve an average error rate of seven to ten percent. This is a reasonable frustration level—high enough to force the full range of reading strategies to become apparent but not high enough to cause a collapse in performance. For our developmental readers in the target grades two, four, and six, these passages were at grade levels four, seven, and ten, respectively. The passages designed for each grade were equated for mean standardized Frequency Index (Carroll, Davies and Richman, 1971), the number of cohesive ties as measured by Fine's (1978) adaptation of the Halliday and Hassen cohesion analysis, and equivalent readability. Readability was based on the Fry (1977) and Spache (1968) methods for grade 4 materials, and the Fry and Dale-Chall (1948) procedures for more difficult passages.

Separating information sources. Leu's third criticism pinpoints a crucial weakness of present methods of error analysis. It is rare for an error to reflect only word based or context based information sources. The substitution of *still* for *silk* or in contrast *frog* for *toad* are relatively pure examples of these rare categories. Typically, errors reflect a mixture of both word and context based information (e.g., *bounced* for *bounded*). Both the *Taxonomy* and the *RMI* deal with this situation by first measuring the predominant information source for each error, then computing percentages of errors in which any given information source predominates. In this way, the role of the minor information sources is lost. There is no method to measure the joint contributions or interactions of word and context based information in either individual errors or total performances.

This difficulty can be traced to the scaling issue inherent in the use of qualitative scales. Once the scales are reorganized to produce data capable of submission to canonical analyses, it will be possible to measure individual or joint contributions of various information sources to error variance.

The Adapted Scales

The initial set of 11 scales may be summarized as follows:

1. Utterance type—incorporates standard error categories: (1) insertion, (2) reversal, (3) substitution, and (4) omission.
2. Word type—measures the degree to which the error substitution is word-like. (Categories are: (1) word, (2) morphologically word-like utterance, (3) accidental esoteric utterance and (4) nonsense or phonemic word-part utterance).
3. Form class of target word (a 21 item test based on semantic adjustments to traditional form class distinctions).
4. Phonemic similarity of the utterance to target word—(proportional scale comparing graphemes in target and utterance based on a grapheme test derived from Venezky).
5. Graphic similarity—(incorporates graphic aspects: inflections, length, and ascenders).
6. Semantic network relationship—measures associative relationship of the utterance to the target word (error is: (1) synonymous, (2) subordinate or superordinate, (3) a general co-hyponym to the target word, or (4) none of the preceding categories applies.
7. Syntactic acceptability of the utterance (an adaptation of Hood's scale of contextual appropriateness).
8. Semantic acceptability of the utterance (as above, but adapted for semantic acceptability).
9. Meaning change (from Goodman and Burke, 1972).
10. Correction—measures the degree to which the attempt brings the utterance closer to the target word.
11. Regression (the number of words re-read).

Reliability. The inter-rater reliability among three coders was determined for a random sample of 500 errors, with each error checked for each of the scales used in this study. Of the sample, 401 errors received identical ratings, and an additional 60 errors were rated within one point difference. For this sample, we have 96 percent agreement within one point and 80 percent exact agreement. Having achieved a reasonable degree of reliability, we undertook an initial test of the system.

METHOD

Subjects

To examine the patterns of reading at three key points in development, 14 students were selected from grade 2, 15 from grade 4, and 22 from grade 6. All students had WISC total IQ scores within the bright normal range and average reading abilities as measured by the *Word Identification* and *Passage Comprehension* measures of the *Woodcock Reading Mastery Test, Form A*.

Procedure

Students were tested individually. Each student read first the narrative then the science passage targeted for his/her grade. Each passage was followed by a cloze comprehension test of 30 deletions. The total testing took an average of thirty minutes including the reading of a randomized alphabet and a random word list designed to gather baseline measures of reading speed. Results for the 51 students were scored subsequently in the eleven scales or audio tapes of the sessions.

Results

Approximately 2600 errors from the three grades were coded. Children in grade 2 averaged 55 errors total for both passages, those in grade 4 averaged 65 errors while sixth graders averaged 42 errors.

A series of factor analyses was performed on the data. Each was a verimax rotation of the principal components solution with equal communalities. The first set of analyses was for the errors generated by each individual, the second set was for all errors generated from a given grade, and the final set included all errors from all grades.

The factor patterns that emerged were highly similar for all the analyses. In each case the first three factors accounted for 65 to 88 percent of the variability, with the lower ranges associated with the individual respondent results.

TABLE 1
Factor Loadings for Eight Scales, Combining Grades Two, Four and Six

	Factors		
	1	2	3
Phonemic Similarity	.007	−.053	.753
Graphic Similarity	.106	−.051	.803
Semantic Network	.725	.089	.316
Syntactic Acceptability	.879	.014	−.113
Semantic Acceptability	.895	.002	−.087
Meaning Change	.818	.035	.209
Correction	−.054	.871	−.103
Regression	.133	.863	−.033

This consistency indicates a similarity of error patterns across subjects and grades and a redundancy in the 8 scales. Since the analyses were highly similar, only one set of weightings (based on the total sample) is presented in Table 1.

These data show that factor 1 loads on semantic network, semantic acceptability, syntactic acceptability and meaning change (a general semantic factor), factor 2 loads on correction and regression (a correction factor), and factor 3 loads on phonemic similarity and graphic similarity (a decoding factor).

The mean factor scores for each subject had non-significant correlations with the verbal and performance subscales of the WISC, suggesting that these measures are not related to general intelligence. Since these results come from students of average IQ's, a similar result may not hold for other groups.

DISCUSSION

These scales are an initial attempt to develop the clinical practice of oral reading error analysis into a standardized procedure which would allow wider generalization of the findings. This initial test of a major part of the system supports its usefulness in several ways.

First, the consistency of the factor patterns over individuals and grades suggest that the system reliably measures some structures in children's strategies as they cope with difficult words. Because the factor structure is so highly similar across grades and subjects, it will be possible to use each of the factor scores to represent its subscales, thus allowing measures of the extent to which readers use word or context based strategies over development.

Additionally, scores for each of the factors may be derived for each error as well as for individual subjects allowing a more valid measure of the joint contributions of information systems to various moments of reading.

The factor structure. Two of the factors found in these data, factors 1 and 3, follow the distinction between data and conceptually driven strategies often made in the literature. The fact that scales drawn from a variety of sources have clustered into these divisions is evidence that our measures provide a useful index of these two types of processing.

Factor 2, the correction factor appears to reflect self-monitoring systems, and probably measures a type of meta-cognitive awareness. Such a factor has not typically been included in reading models. Its inclusion in our system will permit measurement of the importance of successful self-correction to reading development. With further refinement and development of the remaining scales, the procedures will have the potential to measure the roles of conceptually driven, data driven and correction factors using various readers and tasks.

ACKNOWLEDGEMENT

Preparation of this article was supported by NICHHD Grant #HD 12278-03.

REFERENCES

Carroll, J., Davies, P., and Richmond, B. *Word frequency book*. New York: Houghton Mifflin, 1971.

Clark, H. A. The language-as-fixed effect fallacy: A critique of statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335–359.

Dale, E., and Chall, J. S. *A formula for predicting readability*. Columbus: Ohio State University, Bureau of Educational Research, 1948.

Fine, J. Conversation, cohesive and thematic patterning in children's dialogues. *Discourse Processes*, 1978, *1*, 247–266.

Fry, E. B. Fry readability graph: Clarification, validation and extension to level 17. *Journal of Reading*, 1977, *21*, 242–252.

Goodman, K. S. Analysis of oral reading miscues: Applied psycholinguistics. *Reading Research Quarterly*, 1969, *5*, 9–30.

Goodman, K. S. Words and morphemes in reading. In K. S. Goodman and J. T. Fleming (Eds.), *Psycholinguistics and the teaching of reading*. Newark, Delaware: International Reading Association, 1968, 497–508.

Goodman, K. S., and Burke, C. L. *Study of children's behavior while reading orally*. Final report. Office of Education—Bureau of Research, 1969, ERIC DOCUMENT ED 021698.

Goodman, K. S., and Goodman, Y. M. Learning about psycholinguistic processes by analyzing oral reading. *Harvard Educational Review*, 1977, *47*, 319–333.

Goodman, Y. M., and Burke, C. L. *Reading miscue inventory*. New York: Macmillan, 1972.

Haupt, E. J., and Goldsmith, J. S. *Oral reading error coding system, test version*. Reading Disabilities Research Institute, College of Medicine and Dentistry of New Jersey, mimeo, 1980.

Hood, J. Qualitative analyses of oral reading errors: The interjudge reliability of scores. *Reading Research Quarterly*, 1976–77, *11*, 557–598.

Leslie, L. The use of graphic and contextual information by average and below-average readers. *Journal of Reading Behavior*, 1980, *12*, 139–149.

Leu, D. J. Oral reading error analyses: A critical review of research and application. *Reading Research Quarterly*, 1982, *17*, 420–437.

Rumelhart, D. E. Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, New Jersey: Erlbaum, 1977, 573–603.

Spache, G. D. *Good reading for poor readers*. Champaign, Illinois: Garrard, 1968.

Stanovich, K. E. Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 1980, *16*, 32–71.

Venezky, R. *The structure of English orthography*. The Hague: Mouton, 1970.

JOHN W. LOGAN
Texas A & M University

CLASSROOM OBSERVATION SYSTEMS: AN HISTORICAL PERSPECTIVE

The Evolution of Teacher Observation Systems

The scientific study of reading originated in the laboratories of Europe in the mid 1850s with experiments conducted on eye movements and reading. By 1900, only fourteen scientific investigations had been conducted in the United States and England (Gray, 1925). During the time period of 1914 to 1915, emphasis was placed on the use of reading tests in the United States and two-thirds of all reading studies reported at that time were associated with the organization, standardization, and application of reading tests. The development and use of such tests facilitated the examination of pupils' reading achievement under classroom conditions.

Researchers, administrators, and teachers began to investigate the efficacy of reading instruction, compare achievement in oral and silent reading, determine the factors associated with reading progress, examine individual pupils' reading strengths and weaknesses, and measure the progress of pupils taught by different methods (Gray, 1925). The issue of effective reading instruction was also being addressed. However, the observation of instruction in natural settings was not incorporated in these investigations. The only instrument being used to examine classroom teacher behaviors was the *teacher rating form*.

Research efforts to observe and examine classroom teacher behaviors were conducted in 1915 when A. E. Boyce reported findings from a landmark investigation of teacher rating forms employed by various public school systems. His inquiry involved surveying public school systems in 350 major cities to gather information on the methods of determining and recording the efficacy of teachers. The results of this early investigation revealed that about sixty percent of the school systems surveyed used some form of rating or formal evaluation to assess the effectiveness of their teachers. Boyce compiled the similar variables that existed among the rating forms. He reported that the most frequently listed variable was "discipline", followed by "instructional skills" and "cooperation", respectively. It is logical to assume that school systems using these rating forms placed value on using such variables as elements of effective instruction.

A follow-up investigation of the Boyce study was conducted in 1925 by Leroy King for the National Education Association. King also analyzed a sample of the rating instruments currently used by large public school systems in major cities. He reported "technique of instruction", "teacher personality", and "classroom management" as the three most predominantly listed factors in the teacher rating forms examined.

Reavis and Cooper (1945) researched specific items in existing rating forms in a comprehensive report on the evaluation of merit in city school systems. They analyzed over 1500 items noted in eighty-five scales and classified the items into seven broad categories. Their study showed that the most frequently listed variables in the rating scales examined were "social relations", "instructional skills", and "teachers' personal characteristics".

A major portion of their study was devoted to a review of the *value* of rating forms commonly used in school systems. They reported major weaknesses such as lack of definitions of terms, ambiguous terminology, and items that linked two independent variables for a single rater judgment. It was their contention that teacher rating forms neither followed nor accurately sampled a pattern of general teacher ability. This information lent support to earlier investigations by Rugg (1922) and Barr (1920) that found lack of agreement among raters, plus the failure of supervisory ratings to significantly correlate with pupil achievement. On the basis of these reviews Reavis and Cooper (1945) concluded that:

> Ratings appear to be invalid, then, as comprehensive measures of either general or specific teaching ability. There still remains for them, however, the sphere of personal opinion . . . It is apparent that ratings are valid measures of the rater's opinion of the teacher (p. 87).

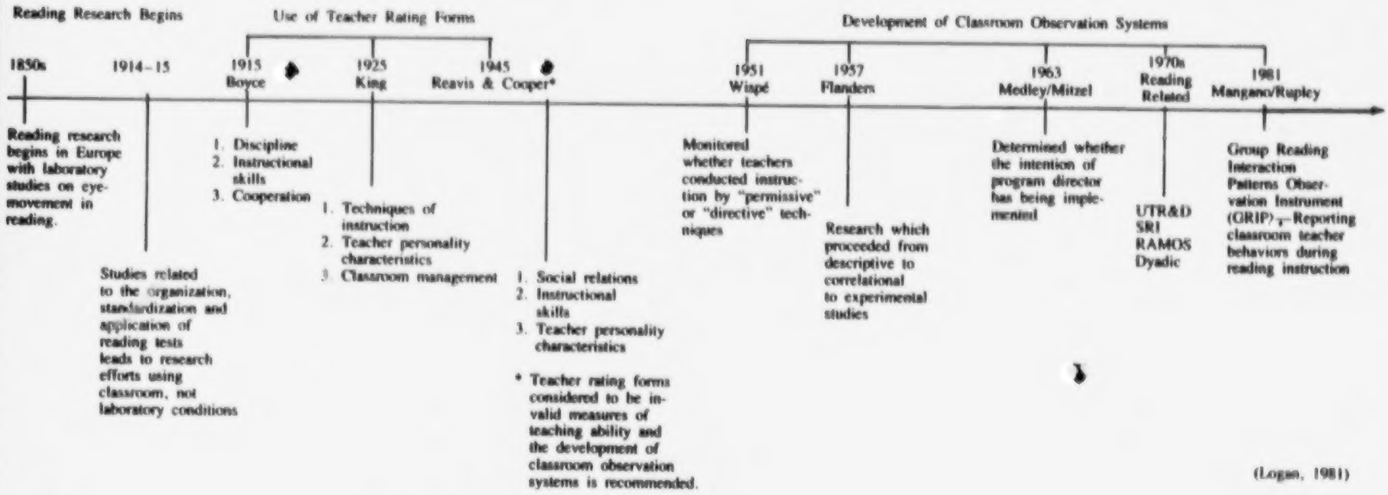
Investigations into teacher rating forms advanced the idea that teacher ratings served mainly as measures of teachers' prestige and that alternative methods needed to be employed to assess features of teaching success. From these investigations came the recommendation of incorporating *observational records of classroom techniques* as a promising new development in order to more effectively gather information on classroom teacher behaviors (Davis, 1964).

Observational Systems Defined

In order to facilitate understanding the purposes of specific observation systems, a general description of classroom observational systems is provided. The purpose of a classroom observation system is to describe, categorize and attempt to measure classroom teacher-pupil behaviors (Simon and Boyer, 1974). Ideally, observation systems represent sets of mutually exclusive behaviors. That is, an observation system should provide a category that represents every behavior observed, and each behavior, in turn, should fit and match a single category. However, in practice the observation system falls short of accomplishing this task in two respects, a category for every behavior observed is not available and many behaviors often seem to fall into two or more categories of the system. In order to compensate for uncategorized behaviors, many systems incorporate a miscellaneous category. Construction of classroom observation systems falls between the categories of 1) the development of a sophisticated system with a large number of categories that provides for fine distinctions, and 2) a system with a few categories that allows for only a gross distinction among behaviors. A system that focuses on several distinct behaviors is more valuable in specifying discrete instructional behaviors of both teacher and pupils; however, such a system requires extensive training of observers. A system that focuses on gross instructional variables is easier for training observers, but may have questionable value in delineating process variables associated with pupils' reading achievement. Many observation systems rest within the boundaries of these two extremes (Simon and Boyer, 1974).

Assessing the Purposes of Classroom Observational Systems

When selecting an observation system to study classroom teacher behaviors, it is insufficient to examine observational category systems only on the basis of the authors' purposes. As systems are adapted for use by other researchers, the original author's purpose often becomes obscured. The result of such adaptations may render the results invalid in relation to the original developer's interest; therefore, it is important to identify the specific uses for which the in-


Reading Research Begins
Use of Teacher Rating Forms
Development of Classroom Observation Systems
1850s
1914–15
1915
Boyce
1925
King
1945
Reavis & Cooper*
1951
Wispé
1957
Flanders
1963
Medley/Mitzel
1970s
Reading
Related
1981
Mangano/Rupley

Reading research begins in Europe with laboratory studies on eye-movement in reading.

Studies related to the organization, standardization and application of reading tests leads to research efforts using classroom, not laboratory conditions

1. Discipline
2. Instructional skills
3. Cooperation

1. Techniques of instruction
2. Teacher personality characteristics
3. Classroom management

1. Social relations
2. Instructional skills
3. Teacher personality characteristics

* Teacher rating forms considered to be invalid measures of teaching ability and the development of classroom observation systems is recommended.

Monitored whether teachers conducted instruction by "permissive" or "directive" techniques

Research which proceeded from descriptive to correlational to experimental studies

Determined whether the intention of program director has being implemented

UTR&D
SRI
RAMOS
Dyadic

Group Reading Interaction Patterns Observation Instrument (GRIP)$_3$—Reporting classroom teacher behaviors during reading instruction

(Logan, 1981)

FIGURE 1. Classroom Observation Systems: An Historical Perspective.

strument had been developed. Information about the uses to which any single system has been incorporated is very difficult to find (Rosenshine and Furst, 1973). Rosenshine and Furst also report that most of the systems developed either describe classroom interaction or provide feedback to teachers. They stress that description alone has not improved the understanding of functional relationships between classroom events and student growth. They also note that only one system in 10 has been used by its author for the purpose of discovering relationships between instructional activities and student growth, and at best, only one system in 6 has been used by anyone in a study involving student-growth measures. In other words, for every six systems that are written, only one author proceeds to do any correlational or experimental study in student growth. Also, for every researcher who uses an existing system for any purpose, five or six others write new, invalidated systems.

Prominent Observation Systems

Many observation systems are available that have been constructed for as many purposes. Several of these systems have been well-established as precursors of present observational investigations. The Flanders' System of Interaction Analysis and the Observation Schedule and Record (OSCAR) are two such notable systems.

One of the earliest attempts to use observational analysis sought to monitor whether teachers conducted instruction by "permissive" or "directive" techniques (Wispé, 1951). However, a major thrust in the development of classroom observation systems occurred in the late 1950s. During this period, N. A. Flanders developed what has expanded to be the Flanders' System of Interaction Analysis, the most widely used classroom observation system. The Flanders' System has been used to provide teachers with a means of obtaining feedback about their own teaching behaviors and the effects of those behaviors on the quantity and quality of student participation in the classroom. Classroom teacher behaviors can be coded by a trained observer sitting in the classroom. The Flanders' contains only ten categories for coding. Despite this, it has proven useful in research and teacher training, partly due to the use of Flanders' simple yet sophisticated matrix technique for coding classroom behaviors (Simon and Boyer, 1974). One of the features that highlights the coded matrix is that it indicates to the reader what preceded and what followed every verbal behavior of both the teacher and the pupils.

The Observation Schedule and Record (OSCAR) was developed to analyze both teacher and pupils' verbal behaviors. The system codes both teacher interchanges and monologues. The interchanges note only teacher behaviors, with emphasis placed on how the teacher initiates an interchange and then how he/she responds to the pupil's answer. By examining a completed coding sheet, the reader should be able to identify immediately whether the teacher responded with explicit or implicit praise or rejection, responded neutrally, or ignored pupil(s) comments. It is recommended that observation periods be at least thirty minute sessions (Simon and Boyer, 1974).

Several notable observation systems have been recently developed that are *specifically* reading related include Teacher and Child Dyadic Interaction Instrument, Coding System for the First Grade Reading Group Study, Secondary Observation Instrument (SRI), Reading and Mathematical Observation System (RAMOS) and Group Reading Interaction Patterns Observation Instrument (GRIP). In most cases, these systems attempt to relate classroom behaviors with pupil achievement through the use of trained observers who code while reading instruction occurs. The value of such systems is that they focus specifically on reading instruction, which enhances the external validity of findings for classroom reading instruction.

Implications for the Future

Classroom reading observation systems offer great potential for the accurate identification and categorization of instructional activities at different levels of specificity. Increased attention to the development of quality reading observation systems will enable researchers to specify patterns of process variables and methods of displaying reading instruction events that are more useful for selection, modification and adoption in studies that cycle through the correlational and experimental process.

REFERENCES

Barr, A. S. *Characteristic difference in the teaching performance of good and poor teachers of the social studies*. Bloomington, Ill.: Public School Publishing Co., 1929.

Biddle, B. J., and Ellena, W. J. (Eds.), *Contemporary research on teacher effectiveness*. New York: Holt, Rinehart and Winston, 1964.

Boyce, A. E. Methods of measuring teachers efficacy. *14th Yearbook National Society for the Study of Education, part 2*. Chicago: University of Chicago Press, 1915.

Davis, Hazel. Evaluating teacher competency. In B. J. Biddle and W. J. Ellena (Eds.), *Contemporary research on teacher effectiveness*. New York: Holt, Rinehart and Winston, 1964.

Gray, W. *Summary of reading research*. Chicago: University of Chicago Press, 1925.

King, L. A. The present status of teacher rating. *American School Board Journal*, 1925, *70*, 44–56.

Mangano, N. G., and Rupley, W. H. *Group reading interaction training manual*. College Station, TX: Instructional Research Laboratory, Department of Educational Curriculum and Instruction, 1982.

Reavis, W. C., and Cooper, D. H. *Evaluation of teacher merit in city school systems*. Supplementary Education Monographs, No. 59. Chicago: University of Chicago, 1945.

Rosenshine, B., and Furst, N. The use of direct observation to study teaching. In Robert M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally, 1973.

Rugg, H. O. Is the rating of human character practicable? *Journal of Educational Psychology*, 1922, *12*, 425–501.

Simon, A., and Boyer, E. G. (Eds.), *Mirrors for behavior, III: An anthology of observation instruments*. Communication Materials Center, Wyncote, Pennsylvania, 1974.

Travers, M. W. (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally, 1973.

Wispé, L. G. Evaluating section teaching methods in an introduction course. *Journal of Educational Research*, 1951, *45*, 161–186.

ROBERT B. RUDDELL
University of California, Berkeley

CHARLES K. KINZER
Peabody College of Vanderbilt University

TEST PREFERENCES AND COMPETENCIES OF FIELD EDUCATORS

Testing plays a major role in interpreting academic achievement of schools to the public and state legislatures throughout the nation. Little information is available, however, to indicate how the classroom teacher, principal, and district office personnel use test data, their technical knowledge of testing which affects interpretation of data, and their perception of the effect of testing on the curriculum in schools (Singer, et. al., 1980). These areas were of central importance in this research. The following research questions served to guide the study:

1. How are tests used by teachers, principals and central office personnel?
2. What technical test/knowledge is possessed by school personnel?
3. What impact does testing have on the school curriculum at the school site and district wide level?

This research was conducted in the state of California, where like other states, schools utilize tests developed by the State Department of Education. The intent of the California Assessment Program (CAP) is to provide information to the legislature and the public on basic skills achievement of California students. The program is designed to compare California students with student achievement nationwide, to assist in decision making in distribution of funds, and to attempt to identify causal factors which contribute to basic skills achievement. To conserve testing time, a multiple matrix sampling procedure is used. With the exception of the grade one entry test no student responds to more than ten percent of the items on any one test, requiring that a composite group test score be derived for a given district. No classroom or individual scores are reported and within the district the minimal unit for reporting is the individual school site. In addition, however, most schools administer a standardized achievement test selected at the district level in order to provide individual and classroom evaluation data.

METHOD

The population for the study consisted of eight school districts selected on the basis of size, location, and student composition. Two districts represented large urban populations with high percentages of minority students identified as low SES based on data provided by the State Department of Education. Two districts were of medium size in a suburban setting representing a middle SES level. One district was of small size, suburban, and of high SES student composition. Finally, one district was of small size, in a rural location at the mid SES level. These districts ranged in location from Southern California to the San Francisco Bay Area and rural Northern California.

Following the formulation of the research questions a 25 item questionnaire and a detailed interview schedule were constructed, field tested, and revised. Combined time for the administration of both instruments consisted of approximately one hour.

Arrangements were made with the central district office in each district for interviews with teachers at the third grade and sixth grade levels, the school principal in each teacher's building, and central office personnel. District office personnel were asked to establish the interviews with school personnel in a school or schools best representing a composite makeup of the district. Questionnaires and interviews were conducted with thirty-eight individuals consisting of 18 teachers, 12 principals, and 8 district office personnel.

Variation in response to test use for personnel within a district exceeded that between districts. An inspection of the data revealed that for analysis and interpretation purposes the data from the eight districts could be pooled for discussion purposes.

RESULTS

The first research question on test utilization produced results which indicated that the most frequent use of tests was for diagnosis, placement, and measuring student growth. These responses were found with greatest frequency for the third and sixth grade teachers, ranging from 33% to 100% and followed by principals and district office personnel, ranging from 17% to 33%. Special program placement for the elementary teacher also revealed a higher frequency, 22% to 44%, than for principals and district office personnel, 0% to 17%. In contrast, principals and district office personnel used tests more frequently to determine trends, 25% and 50%, respectively, than did teachers, 0%.

Information on use of tests to assess instructional programs revealed that teachers and principals make greatest use of tests in "diagnosis" of classroom programs, ranging from 50% to 87%. Principals expressed greatest concern that a close relationship be present between test items and the curriculum, reporting at a 50% level. The use of tests for program evaluation, not pupil assessment, is of greatest concern to district office personnel, 33%, and to a minor degree the principal, 8%. Both district office personnel and principals express the view that test use is important in providing information to the state department of education and the legislature, each at the 17% level.

All school personnel were concerned about specific limitations of tests. The third and sixth grade teachers believed that tests should measure the actual performance of pupils as the reading process occurs, rather than attempt to simulate this process in a test format, 22% and 44%, respectively. Teachers were also concerned that testing not be the sole criterion for measuring progress, ranging from 33% at the third grade to 22% at the sixth grade level. Principals expressed the need for greater availability of subtest interpretation to assist in curriculum planning, 50%, while district office personnel were most concerned that testing consumed too much time because of multiple test administration, 76%.

The second research question dealt with test knowledge possessed by school personnel. All individuals expressed the view that achievement test performance was closely related to reading ability. Test taking skill was identified as contributing to high performance by principals, 33%, and 6th grade teachers, 11%. Most consistent agreement on a single factor affecting high test performance was student attitude toward test taking, with strongest opinions held by third grade teachers, 56%, and principals, 42%. A smaller percentage of teachers believe that test performance is influenced by careful work, 22%, self motivation and confidence, 22%, and parents that value education, 11%. A comparatively high percentage of principals expressed the view that parents who value education influence test results, 42%.

The most important criterion for test selection identified by all personnel was the fit between the test and the curriculum, 58% to 89%. The readability of test items was considered of some importance to third and sixth grade teachers, 22% to 33% respectively, and to principals, 25%. District office personnel expressed greatest concern that the test be easy to administer, 50% and that recent norms be present, 33%.

Test preference responses indicated that teachers favor criterion referenced tests over standardized tests by a 78% margin. The third and sixth grade teachers believe that criterion referenced tests provide a more complete measure of individual growth, 78% to 44%, and again value items requiring performance which closely parallels the reading process present in their instruction, 67% to 44%, respectively.

While the above information provides some clue to test knowledge the most direct indication of technical test knowledge came from one interview item on the interview schedule. Specifically, a brief section of the California Assessment Program informational report used for local school interpretation was provided each individual and interpretation of the standard error of measurement was requested. Of the 38 individuals interviewed correct interpretation was provided by two teachers, one principal, and two district office personnel. Principals and district office personnel expressed the view that teachers and administrators do not understand technical aspects of testing, 42% and 83% respectively.

The third research question explored the effect of testing on the curriculum. While the majority of school personnel expressed the viewpoint that testing has some impact on the curriculum, 33% of the third grade teachers, 89% of the sixth grade teachers, 25% of the principals and 33% of the district office personnel believe that testing has no direct impact on their curriculum. Teachers expressed the view that tests are used to reaffirm teacher evaluation, 33%, with principals at the 17% level. The belief that test results were so general as to provide no help was reported by 33% of the third grade teachers and 44% of the sixth grade teachers.

Specific concern was given to emphasis placed on instructional preparation for achievement testing. A remarkably high percentage of teachers, 33% to 67%, indicated that some form of test taking behavior was taught. This included practicing test taking strategies, 56% to 67%; direct teaching of skills and strategies expected, 44% to 56%; and review of expected questions and formats, 33% to 56%. Conversely, 22% of the third and sixth grade teachers reported that no direct teaching to tests was present. Principals responded at a 33% level and district office personnel at a 67% level that no direct teaching to tests was present.

CONCLUSIONS

Based on the findings of the study three conclusions may be drawn. First, in regard to test use the majority of teachers, principals, and district personnel do believe that testing contributes to their curriculum efforts. Teachers and principals use test data which yield results most closely associated with the individual student for diagnosis, placement, and measurement of growth. District office personnel, and to some degree principals, use test data which serve to identify achievement trends for the school district and state department of education.

Second, in regard to test knowledge there is a general absence of technical knowledge present for the teacher, principal, and curriculum personnel in the central office. There is, however, an awareness at all levels of key factors which are essential to a good test. These range from a good fit between test and curriculum to the importance of test taking strategies and from the importance of recent norming to readability of test items. Teachers overwhelmingly favor criterion referenced tests in contrast to achievement tests because data can be more effectively used in identifying instructional problems with the student. Further, teachers believe that test items should approximate normal reading material and test performance requirements should be as close to normal reading behavior as is possible. Both third and sixth grade teachers express the view that tests should not be the sole criterion for evaluation. Principals state that there is greater need for subtest information to assist interpretation, while district office personnel express concern that testing time is too great and moves teachers and students away from direct instruction. The test sense of school personnel appears to have been derived from classroom and test administration experience rather than from any technical study of testing and evaluation.

Third, the majority of teachers and principals believe that tests have impact on the curriculum. There is strong accountability pressure placed on teachers and principals as well as central office personnel. Perhaps then, it is not surprising to find the high percentage of teachers and principals who feel compelled to teach test taking strategies and skills, and in some cases review expected test questions and test formats. One can only speculate on the impact of such instruction on achievement test results. Perhaps the best advice is found in developing more modest expectations for testing and in confronting the realization that achievement testing provides only one data source for instructional decision making.

ACKNOWLEDGEMENT

This research was supported by the California Policy Seminar, University of California (John Cummins, Director).

REFERENCE

Singer, H., Ruddell, R., McNeil, J., & Whittrock, M. *Testing achievement in basic skills*, California Policy Seminar, University of California, Berkeley, 1980.

WILLIAM H. RUPLEY
NANCY G. MANGANO
Texas A & M University

DEVELOPMENT AND MEASUREMENT ISSUES ASSOCIATED WITH READING CLASSROOM OBSERVATION SYSTEMS

In recent years the study of effective classroom reading instruction has moved from process-product correlational studies to experimental research in natural settings. This new experimental focus has been described by Gage and Giaconia (1981) as having two major stages. The first stage consists of training one group of teachers to employ process variables associated with effective instruction and withholding training from a comparable group of teachers. Identified process variables utilized in the training are measured through the use of observation systems in the teachers' classrooms. The second stage is characterized by using the observation data as the independent variable and students' product outcomes, such as reading achievement, as the dependent variable. Gage and Giaconia state that, "the results of the second stage throw light on the degree to which the new kinds of teacher behavior actually improve a given kind of educational outcome as measured in students" (1980, p. 4).

The results of this new research focus in reading teacher effectiveness research hold considerable promise for specifying instructional process variables that cause students' reading achievement. However, as researchers continue to advance this line of inquiry careful consideration needs to be given to the development of observation systems and the training of observers in the use of such systems in classroom settings. If the observation system used to identify the degree of process implementation in the classroom is not carefully conceptualized and rigorously developed, then, the teacher behaviors associated with student product outcomes will have limited external validity (Mangano, Rupley, & Willson, 1981).

ISSUES RELATED TO OBSERVATION INSTRUMENTS

Researchers who are interested in conducting effective reading instruction inquiries utilizing the above noted two stage model need to address the major issues related to observation instrument development. One concern expressed in the literature is that many instrument developers have failed to either take into account or report the methods that they used to establish reliability. Herbert and Attridge (1975) in an analysis of articles related to observation methodology found only 11 articles from a group of approximately 400 that gave emphasis to reliability data. Furthermore, there is considerable evidence to indicate that most researchers regard observer agreement as the only form of reliability that is necessary to address in the development of classroom observation systems (Frick and Semmel, 1978).

Another criticism aimed at developers of observation systems is that most of them have not given appropriate attention to the generalizability of their findings. Shavelson and Demsey-Attwood (1976) note that the stability of teacher behavior reflects the conceptual framework of generalizability theory. They examined the stability of variables related to teacher behaviors in several studies to determine the adequacy in which one or several observations of a teacher could be generalized over all possible observations. Shavelson and Demsey-Attwood concluded that the stability of measures of teacher effectiveness may be limited in an education context. Although they maintained that the more global variables tended to be stable over time, other variables ranged from moderately stable to unstable.

In a similar analysis, Calkins, Borich, Pascone, and Kugel (1977) described data that were obtained from five observation systems frequently used in process-product research. Reliability and generalizability data were gathered from these systems by five pairs of trained coders. The results indicated that fewer than one-fourth of the behavioral categories in the five instruments was generalizable in any combination of eight or fewer coders with eight or fewer observation occasions.

Major Considerations in the Development of Observation Instruments

The preceding section clearly points out that many developers of observation instruments have been criticized for failure to report or take into account both reliability and generalizability measures. Essentially, careful attention to both of these measures (reliability and generalizability) should be a major phase in observation instrument development and data gathering. Procedures that can be used to attend to both reliability and generalizability of observation instruments are presented in the following sections. These procedures are those that were employed in the two year development of a reading classroom observation system (Mangano & Rupley, 1981).

Reliability Phase

Objective: Determine the reliability of the observation instrument for (1) inter-observer agreement, (2) intra-observer agreement, and (3) criterion-related agreement.

Medley and Mitzel (1963) suggested that measures on observation instruments are reliable "to the extent that the average difference between two measures obtained in the same classroom is smaller than the average difference between two measures obtained in different classrooms" (p. 250). Rowley (1976) argues that an observation instrument is neither reliable nor unreliable, but rather it is when it has been used to collect data and when these data have been manipulated that researchers can speak about its reliability.

Since reliability refers to the consistency of a measure, reliability of an observation instrument is dependent on the degree of accuracy of the observers when using the instrument. This being the case, a sound and extensive training program for observers is essential if reliability measures are to be adequate. Guidelines for developing such training programs are few. In a review of the literature on teacher effectiveness studies few articles were found that summarized the techniques used during the training of observers.

A two year study involving the development of a 58 behavior item classroom reading instruction observation system (Mangano & Rupley, 1981), indicated that a 20 day training period was necessary for reliability coefficients of .80 prior to actual classroom observations. Video-cassettes of teacher and student behaviors were utilized for the training of the observers. This phase of the training provides a measure of inter-observer agreement by pairing each observer with every other observer for each set of coded behaviors in the instrument.

Each pairing is analyzed through the use of simple percentage agreement. The percentage agreement is accomplished by comparing category agreement on individual events and total frequencies of categories across a number of events. Final analysis of agreement is accomplished by using Light's extension of K formula (Light, 1971), which provides a measure of response agreement for each observer across all observations.

Intra-observer agreement, which is a measure of the consistency of each observer with him or herself was addressed by the use of two 15 cassette-video tapes. Each observer is shown the first 15 minute tape during his or her initial training and again early in the actual classroom observation period. Observers code teacher and student behaviors on the second tape after actual classroom observations have begun, usually during the second or third week. Then, toward the conclusion of their actual classroom observations, observers again code the second video-cassette tape. The degree of agreement between each observer's coding of the same tape is determined by using Flanders' π formula (Frick & Semmel, 1978).

Criterion-related agreement was determined within the context of the actual classroom during the beginning, middle, and end of the duration of the classroom observations. A criterion person who is extremely competent with the system is paired with each observer for each of the three periods of the total observation sessions, which in our research ranges from 12 to 30 weeks depending on the purpose of the investigation. Therefore, in a 30 week observation period the criterion person would simultaneously code behaviors with each of the observers during the first, 15th, and 30th weeks of the observation period. The method used to ascertain the agreement between each observer and the criterion person is Light's extension of K formula.

Generalizability Phase

Objective: Determine the generalizability coefficients for each subcategory of the observation instrument.

Generalizability theory assumes that a sample is equivalent to the universe of measures of a particular type (Calkin, Borich, Pascone, & Kugel, 1977). In relation to observation instruments this universe consists of a set of possible combinations of the conditions for which observations can be made. Sets of conditions that vary along the same dimension are called facets. If observations are made within a particular facet, it is then generalizable to other situations that are similar. Generalizability of an observation system is related to facets such as (1) number of raters used to test the system and (2) number of visits for each classroom.

The primary purpose of the generalizability phase was to determine the degree of variablity for each facet. A teacher or instructional behavior is considered generalizable to the extent that scores associated with it are similar across observers and scores for occasions are comparable across occasions. In other words, for a teacher's behavior to be generalizable, it must be (1) consistent from one occasion to another and (2) coded similarly.

A strong case has been made by McGaw, Wardrup and Bunda (1972) that three types of generalizability coefficients must be obtained for observation instruments: (1) generalizability coefficients of reliability for the categories on which teachers are observed, (2) generalizability of the reliability of situations or events, and (3) generalizability with which observers can detect the systematic variation among teachers in their behavior change. Each of these types of generalizability was based on all observational data collected during the study. The procedure that we have utilized is to compute the variance components in the linear models and apply the mean square for each term to the appropriate intraclass coefficient formula, which yields the generalizability coefficients.

CONCLUSION AND FUTURE IMPLICATIONS

As researchers begin to utilize the two stage model for studying reading instruction effectiveness in classroom settings, more careful attention needs to be given to the development of observation instruments that will be used to gather information on process behaviors. If insufficient attention is given to the reliability and generalizability of observation systems, then the results of the study will have limited value in terms of either enhancing the quality of research or the quality of classroom reading instruction.

The development of an observation instrument that is a sound measurement tool is a demanding and time consuming task. However, by addressing the issues of reliabilty and generalizability as outlined above assures a greater likelihood of incorporating sound methodological procedures in the study of teachers and students in natural settings.

REFERENCES

Calkins, D., Borich, G. D., Pascone, M., & Kugel, C. L. Generalizability of teacher behaviors across classroom observation systems. In Borich, G. and others, *Classroom observation data: Is it valid? Is it generalizable?* A compendium of methodological papers, 1977, ED 137371.

Frick, T., & Semmel, M. I. Observer agreement and reliabilities of classroom observational measures. *Review of Educational Research*. 1978, *48*, 157–84.

Gage, N. L., & Giaconia, R. Teaching practices and student achievement: Causal connections. *New York University Education Quarterly*. 1981, Spring, 2–9.

Herbert, J., & Attridge, C. A. A guide for developers and users of observation systems and manuals. *American Educational Research Journal*. 1969, *6*, 227–39.

Mangano, N. G., & Rupley, W. H. Training manual for the group reading instruction process observation system. *Instructional Research Laboratory Technical Paper Series*. Department of Educational Curriculum and Instruction, Texas A & M University, 1981.

Mangano, N. G., Rupley, W. H., & Willson, V. L. *Validity issues associated with observation instruments in teacher effectiveness studies*. Paper presented at the annual meeting of the National Reading Conference, Dallas, December, 1981.

McGaw, B., Wardrup, J. L., & Bunda, M. A. Classroom observation schemes: Where are the errors? *American Educational Research Journal*. 1972, *9*, 13–27.

Medley, D. M. *Teacher competence and teacher effectiveness: A review of process-product research*. Washington D.C.: American Association of Colleges for Teacher Education, August, 1977.

Rowley, G. L. The reliability of observational measures. *American Research Journal*. 1976, *13*, 5–9.

Shavelson, R., & Demsey-Attwood, N. Generalizability measures of teaching behavior. *Review of Educational Research*. 1976, *46*, 553–611.

Stallings, J. A., & Kaskowitz, D. H. *Follow through classroom observation evaluation*. 1972–1973. Menlo Park, California: Stanford Research Institute, 1974.

TIMOTHY SHANAHAN
MICHAEL KAMIL
University of Illinois at Chicago

THE SENSITIVITY OF CLOZE TO PASSAGE ORGANIZATION

Recently, the validity of the cloze test as a measure of reading comprehension has come under close scrutiny. Studies have demonstrated some important limitations of the cloze test, such as insensitivity to background information (Bartoo, 1976; Smith-Burke, Gingrich & Eagleeye, 1978; Tuinman, Blanton & Gray, 1975) or insensitivity to passage organization (Kibby, 1980; Shanahan, Kamil & Tobin, 1982). These studies have failed, however, to provide two rather important pieces of information. First, they have not shown the efficacy of alternative measurement procedures with the same passages, and under similar conditions. Thus, it is not known whether cloze is uniquely insensitive or whether all measures of reading comprehension suffer from the same or similar difficulties. Second, these studies have treated actual reading comprehension or text processing of the subjects as identical to the cloze performance itself. It is impossible, on the basis of such research, to determine whether cloze tests are limited because subjects process information differently in cloze than they might under normal reading conditions, or because the cloze test itself is simply insensitive to subtle differences in comprehension.

Cloze and Intersentential Information Integration

A text includes sentences and propositions that "are related to each other both substantively and by cohesion, and it is characteristic of text that the sequence of sentences cannot be disturbed without destroying or radically altering the meaning" (Halliday & Hassan, 1976, p. 28). Analogously, the task of the reader is often posited to include the integration of the elements of the text into a cohesive and meaningful whole during text processing (Clark, 1977; Frederickson, 1975; Kintsch & van Dijk, 1978). The importance of the process of integration to reading comprehension has been repeatedly demonstrated empirically (Cirillo & Foss, 1980; Frase, 1970; Kintsch, Mandel & Kozminsky, 1977; Kintsch & Vipond, 1978; Stein & Nezworski, 1978; Vipond, 1980).

Given the importance of text integration to the comprehension process, it should be expected that cloze tests would be able to distinguish whether or not it has taken place. To study this, investigators have reduced the possibility of intersentential information integration by scrambling sentence order in one set of passages, and permitting its usage in a second set (Kibby, 1980; Shanahan, Kamil & Tobin, 1982). Average performances on these matched passages have then been compared to find out if the cloze results reflect the manipulation. This research has indicated that, with performance on traditional fifth-word random-deletion cloze tests, it is impossible to distinguish between sequential and scrambled text. This insensitivity is apparent even when test completion time and prior knowledge are held constant, and no matter what the difficulty of the material (as measured by popular readability formulae). The prevention of the beneficial use of any information across sentence boundaries, through the examination of cloze performance on isolated sentences or through the embedding of sentences in non-supportive content, suggests that students may, in certain instances, use information across sentence boundaries. However, such usage appears to be relatively rare and it exerts virtually no influence on the overall cloze test scores.

These studies have successfully demonstrated this important limitation on cloze interpretation, but they have not gone far enough. Granting the importance of intersentential information integration to comprehension and the inability of cloze to measure it, how do we know that multiple-choice questions or retelling measures would do any better given the same passages and conditions? It is possible that, with regard to intersentential integration, cloze does as well as, or better than, other popular measures. Also, this research has failed to indicate whether or not this limitation is due to the way in which subjects process information in a cloze passage, or whether it is an insensitivity of the test itself. This study is designed to test the following predictions:

1) Recall from intact passages was expected to be more complete than recall from the incomplete cloze passages.
2) Recall of information from sequential passages was hypothesized to be more complete than recall from scrambled passages.
3) Scrambling was expected to exert less impact upon cloze recall than upon intact passage recall as previous results indicate that cloze performance is impervious to scrambling.

METHOD

Subjects

Eighty students enrolled in preservice teaching methods courses at a large midwestern university participated in this study.

Materials

A passage of approximately 275 words was selected for use in this study. The passage was a fictional account of an imaginary galaxy, written in an expository style. The experimenters divided the article into 42 discrete idea units according to a method based upon that of Johnson (1970); there were very few disagreements as to the divisions, and consensus was accomplished through discussion. According to the Harris-Jacobson formula (1980), the passage was written at an eighth grade level, indicating that the word recognition demands of the passage should not have prevented any of the subjects from focusing on the understanding of the passage. The sentences were shortened or lengthened as necessary to give all sentences lengths equal to multiples of fives. With sentences of these lengths, it was possible to randomize sentence order and still to delete identical words, at equal intervening distances from both the sequential and scrambled passages. Further details may be found in Shanahan, et al., 1982. This passage was then transformed into four forms. Form I was the passage; Form II was the same passage with the sequence of sentences randomly ordered; Forms III and IV were the same passages presented in a cloze format.

Procedures

Each subject was randomly assigned to one of the four conditions. Subjects were given test booklets which included directions, one passage (I, II, III, or IV), a page of 3-digit multiplication problems, and a page on which to write all information recalled from the passage. Subjects with the intact passages were directed to read the passage twice as if they were studying for an examination.

They were allowed to mark the passage in any way they thought might help them to recall the information. Cloze passage subjects completed the cloze blanks while the others just read. No indication that some of the passages were scrambled was given. When subjects finished reading, cloze completion time or reading time were noted. Students than performed multiplication problems in order to reduce recency effects of short term memory. Finally, all subjects were directed to write down everything that they could remember from the passage without looking back at the passage.

Analysis

Cloze passages were scored using exact replacement criteria. Recall protocols were examined in order to estimate the number of idea units represented from the original text base. No measures of the importance of the ideas to the text hierarchy were used, nor was there any attempt to evaluate the ordering or sequence of units. Performances on the 50-item cloze and the 42-item recall tasks were analyzed in terms of percentage of accuracy. The means and standard deviations for each measure are reported in Table 1.

TABLE 1

Mean Percentage and Standard Deviation of Cloze and Recall Performance and Reading Times, and Standard Deviations, for Four Passage Types

		Cloze (percent correct completions)		**Time (minutes)**		**Recall (percent idea units)**	
Passage Type	**n**	**$\bar{X}$**	**SD**	**$\bar{X}$**	**SD**	**$\bar{X}$**	**SD**
Cloze—sequential	20	48.55	14.37	11.60	3.15	20.30	9.26
Cloze—scrambled	20	46.35	9.74	13.90	4.97	13.35	10.22
Intact—Sequential	20			6.85	4.62	31.10	13.77
Intact—Scrambled	20			7.20	3.32	23.95	10.03

RESULTS

In order to determine whether this study replicated previous findings, a two-sample t-test was calculated to compare average cloze test performances across sentence order conditions. This analysis indicated that there was no significant difference in cloze performance due to sentence order ($t=.05$, $df=39$, $p<.40$).

The regression of recall scores (proportion of idea units recalled) on orthogonal vectors representing sentence order, passage type, and their interaction explained 26.47% of the recall variance ($R=.5145$). Because of planned comparisons of orthogonally coded vectors, the effects of the independent variables are not correlated, and therefore, they are individually interpretable (Kerlinger & Pedhazur, 1973). Passage type ($R^2=.1846$, $F(1,78)=17.65$, $p<.0005$), and sentence order ($R^2=.0849$, $F(1,77)=8.39$, $p<.005$) both accounted for significant amounts of recall variance. An examination of the mean performances of each group (see Table) indicates that cloze passages led to lower levels of recall than did intact passages, and that scrambled passages resulted in lower levels of recall than did the sequential passages. The interaction statement did not contribute significantly to the regression ($R^2=.00002$, $F(1,76)=.002$, $p>.90$).

That is, the effects of passage type and sentence order are essentially additive.

A similar analysis was run with number of minutes subjects read or completed cloze blanks entered as a covariate. This change increased the variance explanation by about 6.5% for the overall regression of recall scores ($R = .57461$). However, sentence order and passage type maintained their relative importance to the regression and the interaction of these variables was again found to be insignificant.

CONCLUSIONS

The results of this study confirm earlier findings that the cloze test is insensitive to the integration of information across sentence boundaries. As in earlier studies, it was found that on the basis of cloze test performance alone it was impossible to distinguish whether subjects had completed a scrambled ordered or sequentially ordered passage. However, free recall protocols did not evidence this deficiency. Passage type was found to be significantly related to level of recall, even with influence of time on task accounted for.

It is important to note that sentence order was related to the recall performance of subjects who completed both intact and cloze passages. Unlike the cloze scores, the recall scores of these same subjects were negatively influenced by the scrambling transformation. Thus, it is not that students fail to integrate information across sentences when completing a cloze exercise, but only that this information does not aid in the sentence completion activity.

This study is not without limitations, however. First, only a single passage was used in this experiment. It is possible that the use of non-fiction material, or narrative material, might have led to different results. In previous investigations of the sensitivity of cloze to sentence integration, a variety of materials have been used, with similar results. Second, the passages used here were quite difficult (i.e., completions averaged less than 50% correct). With passages this difficult it is possible that normal sentence integration is prevented. This argument does not appear to be a strong one as a range of passage difficulties (28–60%) has been used in previous research (Shanahan, Kamil & Tobin, 1982), with similar results. If passage difficulty limited or prevented integration it is doubtful that recall differences would have been found across passage types.

The popularity of cloze as a test of reading comprehension seems to be based, at least in part, upon the moderate to high intercorrelations which exist between cloze and other measures of comprehension (Bickley, Ellington & Bickley, 1970; Rankin, 1965, 1974). The results of this study do not challenge the overlap of cloze with other tests. In fact, the correlation of cloze with recall in this study was .63, indicating about 40% shared variance for these measures. However, this research does suggest that cloze and recall tests weight the importance of at least one component process of comprehension, intersentential information integration, differently. Cloze gives far less emphasis to this dimension of comprehension than does unaided free recall.

If contemporary theories of comprehension, which give great weight to intersentential information integration, are valid, then cloze is not appropriate as a measure of comprehension. It is possible, however, that through some basic design or interpretation change, such as the deletion or scoring of specific cohesive ties or links at certain distances, cloze could be made more sensitive to information integration. Future research needs to evaluate the possibility of other deletion patterns or scoring procedures providing a more adequate measure of intersentential information integration.

REFERENCES

Bartoo, R. K. The effect of cumulative context upon cloze task performance of tenth graders. (Doctoral dissertation, State University of New York, 1975). *Dissertation Abstracts International*, 1976, *36:* 842A.

Bickley, A. C., Ellington, B. J., & Bickley, R. T. The cloze procedure: A conspectus. *Journal of Reading Behavior*, 1970, *2*, 232–249.

Cirillo, R. K., & Foss, D. J. Text structure and reading time for sentences. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 96–109.

Clark, H. Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

Frase, L. T. Influence of sentence order and amount of higher level text processing upon reproductive and productive memory. *American Educational Research Journal*, 1970, *7*, 307–319.

Frederickson, C. H. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology*, 1975, *7*, 371–458.

Halliday, M. A. K., & Hassan, R. *Cohesion in English*. London: Longman, 1976.

Harris, A. J., & Jacobson, J. D. The Harris-Jacobson readability formulas. In A. J. Harris & E. R. Sipay, *How to increase reading ability*, 7th ed. New York: Longman, 1980.

Johnson, R. E. Recall of prose as a function of the structural importance of the linguistic subunits. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 12–20.

Kerlinger, F. N., & Pedhazur, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston, 1973.

Kibby, M. W. Intersentential processes in reading comprehension. *Journal of Reading Behavior*, 1980, *12*, 299–312.

Kintsch, W., Mandel, T. S., & Kozminsky, E. Summarizing scrambled stories. *Memory & Cognition*, 1977, *5*, 547–552.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. *Psychological Review*, 1978, *85*, 363–394.

Kintsch, W., & Vipond, D. Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Memory: Processes and Problems*. Hillsdale, N.J.: Erlbaum, 1978.

Rankin, E. F. The cloze procedure—its validity and utility. In E. L. Thurston & L. E. Hafner (Eds.), *The philosophical and sociological bases of reading*. Fourteenth Yearbook of the National Reading Conference, 1965.

Rankin, E. F. The cloze procedure revisited. In P. L. Nacke (Ed.), *Interaction: Research and practices in college adult reading*. Twenty-Third Yearbook of the National Reading Conference, 1974.

Shanahan, T., Kamil, M. L., & Tobin, A. W. Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, 1982, *17*, 229–255.

Smith-Burke, M., Gingrich, P. S., & Eagleeye, D. Differential effects of prior context, style and deletion pattern on cloze comprehension. In P. D. Pearson & J. Hansen (Eds.), *Reading: Disciplined inquiry in process and practice*. Twenty-Seventh Yearbook of the National Reading Conference, 1978.

Stein, N. L., & Nezworski, T. The effects of organization and instructional set on story memory. *Discourse Processes*, 1978, *1*, 177–193.

Tuinman, J. J., Blanton, W. E., & Gray, G. A note on cloze as a measure of comprehension. *Journal of Psychology*, 1975, *30*, 415–433.

Vipond, D. Micro- and macro-processes in text comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 276–296.

HARRY SINGER
University of California, Riverside

MARIAM JEAN DREHER
University of Maryland, College Park

PREFERENCE FOR TESTING AND REPORTING OF TEST RESULTS IN BASIC SKILLS: ACQUISITION VS. ABILITY TO GAIN INFORMATION FROM TEXT, CONTENT-BOUND AND CONTENT-FREE

Acquisition and use of basic skills can be confounded when standardized tests, norm-referenced for each grade level, are administered to a group of students as they progress through school. For example, improvement in reading acquisition and use of reading for gaining information from text are both reflected in test scores at successive grade levels. These scores are frequently interpreted as percentile rankings. This type of interpretation of test scores only informs consumers on the relative standing of students, not on progress made since administration of the previous test.

To assess any kind of acquisition or learning of a skill, it is necessary to keep constant the task or the goal to be learned. Then, for learning to occur, there must be active reaction of the learner to relevant instructional situations with reinforcement or knowledge of results following the reaction or at least some contiguity observed by the learner between the stimulus and the appropriateness of the response (Hilgrad and Bower, 1966). An assessment task which is appropriate provides knowledge of the results. Based on feedback from the assessment, further instruction with some modification may ensue. Then the same assessment task is readministered and progress noted. In short to assess learning, the paradigm consists of teaching, testing, feeding back and acting on test results, reteaching, and then testing again on the *same task*. Through such a procedure, we can diagnose acquisition of a basic skill from its application.

Differentiating the Basic Skill of Reading

Although learning to read cannot be completely separated from use of reading to gain information in the instructional paradigm, we can nevertheless approximate assessment of learning to read. We can do so by repeating the same reading comprehension test after each period of instruction until students attain, say, a criterion of 85 percent correct comprehension responses or at least 85th percentile ranking on the norms for the readministered test. When they have attained this criterion we can infer that they have made progress in learning (a) correct responses to printed words, (b) use of their language background, knowledge, and reasoning ability in word identification and in comprehension and (c) integration of (a) and (b). We call this type of assessment *repeated* or *self-comparison* testing. We can continue this type of testing until students are able to read and comprehend materials that represent a consensual criterion that the students had learned to read. Emphasis in instruction and assessment then can shift to ability to gain information from the text.

When students improve in use of reading to gain information as reflected in reading achievement on tests standardized and norm-referenced for each grade, we can infer that they are not only gaining in reading acquisition but also in their own resources for interacting with text. These resources include improvement in language ability, background knowledge, and reasoning processes. We call this type of assessment *peer-comparison* testing.

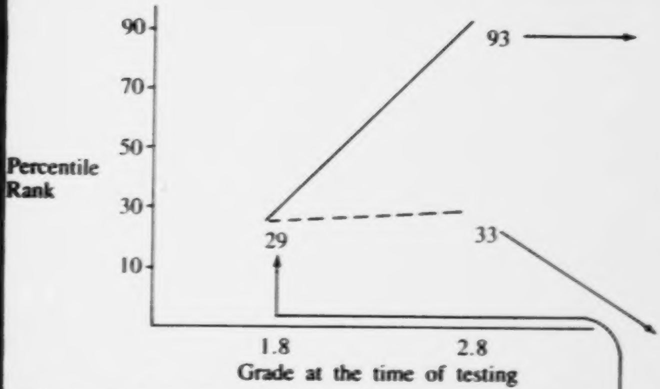
We can contrast self-comparison testing with typical peer-comparison testing. Elsewhere we have used a hypothetical distribution of results to depict what we might expect to obtain by testing a representative class of students on both types of testing (Singer, 1977a, b). The distributions for repeated testing show in this example *decreasing* variability while peer-comparison testing reveals *increasing* variability as a group of students progress through the grades. We can infer from the repeated testing results that the class was developing rapidly in acquisition of reading ability, and becoming more alike in this phase of reading. Also, we could infer that the group was becoming more heterogeneous on the peer-comparison measures of reading attainment. In general, the variability for such peer comparisons is equal to two-thirds of the group's median age. (For an actual distribution of reading achievement for grades 2–8, see Cook, 1951.) Thus, we can differentiate reading acquisition from ability to use reading in interaction with text to attain comprehension for increasingly more difficult tasks.

Content-Free vs. Content-Bound Test Results

Test results, if reported in the usual way, by reporting only percentile ranking, do not inform us on what the students know and can do at any point in the testing program. When releasing test results to the newspapers, when sending test results home, and frequently when reporting test results to pupils, school personnel, and board members, schools typically report only numerical scores, such as percentiles or stanines (Green, 1975). We call such reporting of test results as *content-free*. To provide *content-bound information*, schools would also have to supply information on the content from which the results had been obtained. For example, the schools might report to the press and the public the most difficult paragraph that students had to read successfully to attain at least a 50th percentile ranking on each of the peer comparison tests at successive grade levels, or they could provide parents with the most difficult paragraph that their children had read correctly. If school staff would report test results this way, then students, their parents, and other school personnel would know not only the students' scores but also the content from which the students had successfully read and gained information.

Up to this point, we have explained the rationale and the procedure for differentiating acquisition of reading-ability on a *fixed measure* from ability to interact with text on *increasingly more difficult measures*. We also contrasted content-free with content-bound reports. Thus, we have discussed two types of testing (self-comparison and peer-comparison) and two ways of reporting test results (content-bound and content-free).

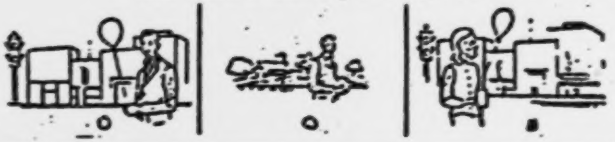
We then used these four ways of testing and reporting test results for depicting a second grade student's test results. All four ways are shown in Figure 1. The first way of testing and reporting, *peer comparison*, the dashed line in Figure 1, shows only the results for the standardized, norm referenced tests administered at the ends of grade 1.8 and 2.8. This type of testing and reporting is the usual way schools test and report test results. The second way included the peer group comparisons *and self-comparison testing*. The self-comparison, as shown by the solid line in Figure 1, indicates the results of the repeated testing using the same test and the same norms for the grade in which the test was originally administered. We repeated these two ways of reporting test information with the addition of *content information*, as shown on the side of Figure 1.



Content-Bound Information

This is the hardest test passage the student read correctly on the first-grade test at the end of second grade. (Note: He got all the passages correct.)

Mrs. Brown is in the city with a balloon.



○ ○ ●

This is the hardest test passage the student could read correctly on the second-grade test at the end of second grade.

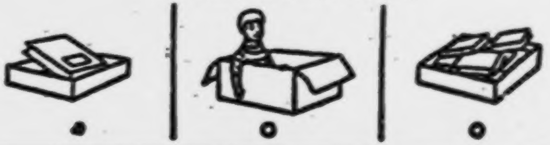
It was Ann's birthday. Mr. Parks had a surprise for Ann. "Here is the box of candy," he said. "Thank you, Father," said Ann, "I will put it on my toy box. I must go to school now. Betty Smith is waiting for me."

Frisky barked when he saw Ann go away. Then he ran to the toy box and got the candy. Frisky had a good time.

What did Frisky do?	What was Father's last name?
○ ran away	○ Brown
○ ate the candy	○ Frisky
○ went to school	○ Smith

This is the hardest test passage the student was able to read correctly on the first-grade test at the end of first grade.

This is one book. It is in a box.



● ○ ○

This graph shows the student's percentile rank on the reading comprehension subtest of the Comprehensive Test of the Basic Skills. At the end of first grade (1.8) this student was tested using the first-grade test. He scored at the 29th percentile, getting 12 out of 24 items correct. At the end of second grade (2.8) the student was given two tests. *First*, he was retested with the same test he had taken in first grade. This retest shows that when compared to students in first grade, he now scores at the 93rd percentile, getting 24 out of 24 items correct. *Second*, he was also tested with the second-grade test. He scored at the 33rd percentile, getting 11 out of 18 items correct. These data were presented in four ways: (a) only peer group comparisons, the dashed line only; (b) the self-comparison and peer group data, both lines as shown in the figure; (c) and (d) the same data as shown in (a) and (b), but with content-bound information, the most difficult paragraph the student could correctly read in each test.

FIGURE 1

BEST COPY AVAILABLE

The content information indicates the most difficult paragraph successfully read in each test.

Purpose

We then wanted to know whether experienced school personnel would prefer one type of testing and reporting of test results over another. We constructed an experimental design to answer this question.

METHOD

Experienced school personnel ranked the four categories of test results, as shown in Figure 1, from most to least preferred.[1] Although order of presentation of the categories probably would not have influenced the rankings, they were nevertheless presented in a random order to each person.

Subjects

A total of 26 subjects participated in the University of California study. They were students enrolled in an administration credential course who had an average of 5.1 years' teaching experience; eight taught at the elementary, four at the junior high, and five at the high school level. Five were administrators at the elementary or high school levels and two had some administrative position. The modal number of courses in tests and measurements taken by the group was one. Two subjects did not complete the information inventory.

The study was repeated at the University of Maryland on 18 students enrolled in a graduate course in research in reading. Thirteen of the students had seven or more years of teaching experience; the others all had from four to six years of teaching experience. Nine were currently teaching at the preschool through elementary level, two at the junior high, and one at the high school level. Two were principals of elementary schools and one had another administrative position. Three subjects neglected to indicate their current positions but were known to be experienced teachers. The modal number of courses in tests and measurements taken by the group was two.

RESULTS

Rank Order

The results of the rank ordering indicated that fifteen subjects preferred repeated testing, content-bound. Their least preferred choice was peer group comparisons only, content-free. The other subjects chose a variety of combinations.

The significance of the choices was evaluated by comparing them with chance selections. The four choices, taken two at a time, can mathematically result in 12 permutations. Hence, each pair of choices has a chance level of one-twelfth. Therefore, only two subjects would be expected to select the same choices by change ($1/12 \times 26 = 2.2$). Consequently, selection of the same pair of choices by 15 subjects is ***seven times greater than chance***. All the other choices were close to, at, or below the level of chance.

This study was replicated on 18 educators enrolled at the University of Maryland, College Park. Twelve of them also had the same choice combinations as those in the first study. This number is eight times greater than chance. The other choices were again near, at, or below chance.

BEST COPY AVAILABLE

DISCUSSION

We know that the National Education Assessment Program provides content-bound information whenever it reports test results, usually in the form of percent of an age group that passes a particular item. In short, reporting results with content-bound information does not involve new tests nor new technology. What is new is use of repeated testing after an interval of instruction and using the same norms for interpreting the results. As a result of this way of testing and reporting test results we *now* know that professional educators would prefer such testing and test reporting, even though it involves adding repeated testing to the testing schedule.

In view of this evidence on educators' preference for diagnostic testing with content-bound information, we are curious as to why schools persist in testing for and reporting only peer group comparisons, content-free. In the absence of evidence, we can only speculate. We recognize that more costs in test administration and scoring are involved, but they are not prohibitive. The content-bound feature can be readily programmed into computers at some relatively small initial cost and printed out in reports with little additional cost.

Copyright on tests might be a problem. However, the Truth in Testing law has mandated release of test item information. In content-bound reporting only a few items would be released to any one student. However, if copyright or test control still constitutes a problem, then comparable items from previous editions of a test might be used with the appropriate qualification that the item is similar to the one the student had taken.

Possibly teachers might object to repeated testing because it entails more testing. But, the teachers who obtained the repeated test data on their classes for this study were delighted because the results were so gratifying to them. Hence, we think that teachers would welcome more testing when the results are likely to be favorable.

However, when results are returned to teachers one to three months after students have been tested, as often happens with machine-scored tests, they are likely to be perceived as irrelevant. The remedy for this problem is to have the teachers or their aides score tests and record the scores before sending the tests off for machine scoring.

CONCLUSION

The educators who participated in this study least preferred testing and test results on peer group comparisons alone, the current way of reporting test results, but most preferred repeated testing and peer-group comparisons with content-bound. This latter type of testing and reporting of test results would reveal to parents and pupils that rapid progress in learning to read occurs, particularly in the primary grades. It would also probably educate the public that schools in general do succeed in teaching most students how to read and how to increase their ability in comprehension or ability to learn from text.

However, the test results will also inform the public that schools do not eliminate individual differences in ability to gain information from text. On the contrary, successful schools help increase this range (Cook, 1951). Perhaps differentiated testing and content-bound reporting will enable the public to make the distinction that almost all students can successfully attain a criterion level in learning how to read, but all students cannot achieve the same criterion level in ability to learn from text. Indeed students can become alike in learning how to read, but they become more and more dissimilar in their ability to learn from text as they progress through school. Whether schools will be able to accomplish this differentiated understanding in pupils, parents, professionals, policymakers, and the press, remains to be seen.

NOTE

1. In a preliminary study, we had students rank order six categories: three types of testing (peer comparison, repeated testing, and both) with content-bound and with content-free. Subjects overwhelmingly preferred the combination testing, content-bound. Consequently in this study, we reduced the categories from 6 to 4 by eliminating the repeated testing alone report. We kept the peer group comparison alone, content-free, type of report because schools typically use it when reporting test results.

ACKNOWLEDGEMENTS

This study was supported by funds from a California Policy Seminar Grant, administered by the Institute of Governmental Studies (John Cummins, Director). The grant was awarded to Harry Singer, John McNeil, Robert Ruddell, and Merl Wittrock, University of California (Riverside, Los Angeles, Berkeley, and Los Angeles, respectively) to study "Achievement Testing in Basic Skills."

For participating in the study, we wish to express appreciation to Mr. Robert Hocker, Acting Superintendent, and the principal, teachers, and students at La Granada Elementary School, Alvord Unified School District; Mr. Ray Berry, Lecturer in Education and his students at the University of California, Riverside; and the students at the University of Maryland, College Park.

REFERENCES

Cook, W. W. The functions of measurement in the facilitation of learning. In E. F. Lindquist (Ed.), *Educational Measurement*. Washington, D.C.: American Council on Education, 1951.

Green, J. Reporting test results in New York City. *National Elementary Principal*, 1975, *54*(6), 91–92.

Hilgard, E. R., & Bower, G. H. *Theories of learning*. New York: Appleton-Century-Crofts, 1966.

Singer, H. IQ is and is not related to reading. In S. Wanat (Ed.), *Issues in Evaluating Reading*. Arlington, Virginia: Center for Applied Linguistics, 1977a.

Singer, H. Measurement of early reading ability. *Contemporary Education*, 1977b, *48*(3), 145–150.

JOHN McNEIL
University of California, Los Angeles

LISBETH DONANT
Atascadero Public Schools

SUMMARIZATION STRATEGY FOR IMPROVING READING COMPREHENSION

The focus of this study stems from a demonstration that the comprehension scores of sixth graders doubled when pupils generated sentences about paragraphs and inserted relevant paragraph headings in their sentences (Doctorow, Wittrock, and Marks, 1978). This demonstration led us to ask and seek an answer to this question: Will training in the rules of summary writing result in better summaries and greater reading comprehension?

Brown, Campione, and Day (1981) identified six basic rules essential to summarization. These rules are similar to the operations described by Kintsch and van Dijk (1978) as basic operations in comprehending and remembering prose. The six rules are (1) deleting unnecessary or trivial material; (2) deleting material that is important but redundant; (3) substituting a superordinate term for a list of items; (4) substituting a superordinate term for components of an action; (5) selecting a topic sentence; and (6) when there is no topic sentence, inventing one. Brown et al. (1981) found that the summaries produced by students from grades 5, 7, 10, and college show developmental differences. Students become increasingly adept at using superordination and selecting topic sentences. However, invention was rarely used by fifth graders and by college students on only half of the occasions when it was appropriate.

Day (1980) has been successful in teaching junior college students to apply the rules. The question remains, however, whether or not young children can be taught the rules and with what effect on reading comprehension?

Subjects

Twenty-three fifth grade pupils in a middle class school were randomly assigned to one of three groups: a summary rule training group, a summary writing group, and a non-instructional control group.

PROCEDURES AND MATERIALS

Pretest

All students were given a pretest consisting of two separate passages. Questions posed about the passages required five inferences. There was no significant difference among pretest scores for the three groups.

Rule Training

Students in the rule training group were given group training on each of the six rules. The reading level of these materials was approximately two grade levels lower than the grade level of the students.

The basic instructional exercises for this training consisted of instructions for twelve passages, such as the following:

"Put a line through all the statements that are *not* important to the story."

> What kind of things make people feel nice? It's nice to *get* something. There couldn't be a better feeling. Or could there? Could it be better to *give* something? Think about it. Some people wish they could take a trip.

"Put a line through all the statements that repeat what has already been said."

> Peter told the cat about the wolves that lived in the forest beyond a meadow. They were large and very, very mean. They were bigger than any dog and mean.

"What is this story about?"

> Sometimes a sentence has marks like these around it (" ") and an end mark like this: (!). Sometimes there is an end mark like this (?). Can you tell how the speaker feels who says, "He found my bike!" and how another speaker feels when he says, "He found my bike?"

"What is this story about?"

> Pete got out the paper. Bill found the paint.
> "Let's draw our street," said Andy. Soon there was a drawing.
> "There's just one more thing," said Peg.
> "See there's no sun in the drawing."
> "Let me fix it."

"Put a line under the topic sentence for this paragraph."

> Mirrors help you in many ways. They help you see yourself. The mirror in a car helps the driver see behind. A dentist might use a mirror to look at your teeth.

"Make up a topic sentence for this paragraph."

> We talk about the *kind* of seeds they plant. We talk about the *kind* lady. We laugh about the *funny* clown and we say it's *funny* when something strange happens. We can *turn* on our toes and we can *turn* on the light. Don't be *mean*—tell me what these words *mean*.

The first selection in each exercise was used to assess the pupil's status with respect to ability to apply the rule. If children could not apply the rule, the teacher modeled it and gave additional practice exercises. For example, if the pupils could not identify irrelevant parts, they were given such classification problems as, "*table, chair, banana*—Which one does not belong?" These problems were followed with activities by which pupils (a) listened to stories and raised their hands when hearing something that did not fit; (b) read paragraphs silently and drew lines under the unimportant parts. After these practice exercises, pupils were asked to apply the rule to the second selection.

Rules were taught in a successive order. Approximately 25 minutes daily were spent on each rule. An exception was in the amount of time given to Rule 5, finding a topic sentence. Students required less time for this rule because it was already familiar to them. Once all rules were taught in isolation, the teacher presented the training group with additional paragraphs taken from classroom materials and asked the pupils to apply each rule and then write a summary for the respective paragraphs. Summaries were not to exceed ten words and not to contain unimportant details. Meanwhile, students in the two control groups carried out their daily classroom activities without special treatment.

Post-test

The post-test consisted of two selections representative of the reading passages found in the California Assessment Program for Assessing Reading Comprehension (California State Department of Education, 1980). The twelve corresponding questions called for the inferential skills for recognizing *main idea, cause and effect,* and *drawing conclusions.* Skills of noting details, recognizing pronoun referents, and sequence were also applicable.

Prior to administering the comprehension questions, the teacher asked those in the training group to apply the six rules to the selections and to write a summary for them.

Students in the summary only group were given a rule card listing the six rules and asked to apply them to the selections after which they were to write their summaries.

Students in the non-instructional control group were asked to read the selections and to be prepared for subsequent questioning.

The selections and summaries were collected and the written questions were given for individually written responses.

RESULTS

Table 1 below shows the percentages of pre- and post-test scores correctly answered by pupils within the three conditions.

TABLE 1
Mean Number and Percentage of Comprehension Questions Correctly Answered by Pupils Under Three Conditions (post-test)

	Training in Rules For Writing Summaries	Writing Summaries Without Training in Rules	Non-Instructional Control
Mean	10.2	7	9.2
Percentage	85%	58%	76%
	(N=8)	(N=7)	(N=8)

Those receiving both training and opportunity to write summaries for the material excelled over both those who wrote summaries without rule training and those who did not receive training or opportunity to write summaries.

Analysis of the summaries written by those in the training and non-training groups showed that only one pupil out of the eight with training failed to write a summary that met criteria of relevance and brevity; while only one of the seven without rule training met both of these criteria.

CS228

DISCUSSION

Students given training in the rules of summary writing appear to write better summaries and achieve higher scores on a reading comprehension test than students who are not given such training. This is true at least for the conditions of the tryout and with recognition that evidence was not collected regarding the degree to which pupils have generated the strategy to the point of using it on their own, applying it to similar learning situations. Asking students to write summaries according to given rules without first training them in the meaning of the rules appears to be detrimental.

Currently, the study is being replicated in eight other schools. Teachers in these schools say that information about pupils' ability to summarize and apply summarization rules is new and important to them. Only one of the six rules—identifying topic sentences—is commonly taught in the present curriculum. These teachers believe that the six skills will contribute to improving comprehension. They indicate that they will continue to teach the rules although two of the teachers said they did not want to generate additional practice exercises, preferring that all exercises be available in textbooks.

CONCLUSION

Research in reading comprehension is bringing new dimensions to understanding the process by which one comprehends text. This knowledge implies that we should change in the ways we teach and assess reading.

Present achievement tests do not provide the teacher with the information needed for helping pupils improve in their reading comprehension. For example, we do not need so many tests to see if the reader can find the main idea, draw conclusions, and determine cause and effect. Instead we need tests to reveal which of the essential processes or strategies for finding the main idea are known and used by the reader.

Similarly, the materials for teaching reading comprehension should do more than offer practice in finding main ideas, drawing conclusions and determining cause and effect. They should provide the uninitiated with techniques for achieving these tasks.

ACKNOWLEDGEMENT

This paper was funded by a California Policy Seminar grant. The grant is administered by the Institute of Governmental Services under the direction of Mr. John Cummins. The grant is to Harry Singer (University of California, Riverside), John McNeil and Merl Wittrock (University of California, Los Angeles), and Robert Ruddell (University of California, Berkeley) to study "Achievement in Testing Basic Skills."

REFERENCES

Brown, A. L., Campione, J. C., and Day, J. D. "Learning to Learn: On Training Students to Learn from Texts." *Educational Researcher*, Feb. 1981, 14–21.

California State Department of Education. *Student Achievement in California Schools*. 1979–80 Annual Report, Sacramento, 1980.

Day, J. D. Training summarization skills: A comparison of teaching methods. Unpublished doctoral dissertation. University of Illinois, 1980.

Doctorow, M. J., Wittrock, M. C., and Marks, C. B. "Generative Processes in Reading Comprehension." *Journal of Educational Psychology*, 1978, *70*, 109–118.

Kintsch, W., and van Dijk, T. A. "Toward a Model of Text Comprehension and Production." *Psychological Review*, 1978, *85*(5), 363–394.

HILDA BORKO
Virginia Tech

TEACHERS' DECISION POLICIES ABOUT GROUPING STUDENTS FOR READING INSTRUCTION

INTRODUCTION

In a recent study of elementary school classrooms, Clements and Evertson (1982) noted that 40 of the 41 teachers in their sample used small group instruction, and that small group instruction was used during reading programs in 85% of their observations. Given this prevalence of small group instruction, one important question that educational researchers can ask is, "What are the factors that guide teachers' decisions in the formation of instructional groups?" The purpose of this study is to examine the strategies used by teachers in forming reading groups. This research is part of a larger investigation of teachers' decisions in the planning of reading instruction.

The research methodology employed in this study is a "policy-capturing" approach in which a linear model is constructed to represent or "capture" the human judgment or decision process (cf. Shulman & Elstein, 1975; Slovic & Lichtenstein, 1971; Slovic, Fischhoff & Lichtenstein, 1977). Use of this methodology in research on teaching is based on a conceptualization of teachers as active agents who make educational decisions by integrating information about students, the subject matter, availability of instructional materials, and other relevant aspects of the teaching situation (cf. Shavelson & Stern, 1981). As one strategy for determining how information is being integrated, teachers are presented with several vignettes (e.g., descriptions of students) constructed by systematically varying information along several different dimensions and are asked to make one or more instructional decisions for each vignette. By regressing the decision on the informational variables, the set of weights or regression coefficients associated with the informational variables can be estimated. These weights are interpreted as indicators of the importance of each variable in the judgment/decision process.

The policy-capturing methodology has been employed to investigate teachers' decisions in a variety of areas including classroom management (Borko & Cadwell, 1982), curriculum content (Floden et al., 1981), educational materials (Yinger & Clark, 1982), and reading and math instruction (Russo, 1978). For example, Russo (1978) examined the effects of student characteristics, teachers' educational beliefs, and the nature of the instructional task on teachers' decisions about grouping students and planning reading instruction. She found that when estimating students' mastery of reading curriculum, teachers relied heavily on standardized reading test scores, while virtually ignoring other student characteristics such as class participation and problematic behavior. Further, about half of the teachers in her study placed students in reading groups solely on the basis of information about reading ability while the remainder took into account math ability as well.

The present study represents a logical next step in the line of research begun by Russo, by incorporating several substantive and methodological modifications. For example, Russo examined teachers' global views about education rather than specific attitudes toward reading and found no significant differences in decision policies that could be accounted for by these views. The present investigation utilizes an instrument designed to measure conceptions of reading (Duffy and Metheny, 1979), which have been shown to guide teachers' instructional practices (Bawden et al., 1979). The present study also examines the role of professional experience, by including both teachers and student teachers as participants.

While the policy capturing approach yields models that predict teachers' decisions, these models do not necessarily reflect the cognitive processes which teachers use in order to arrive at their decisions (Anderson & Shanteau, 1977). The major methodological modification to Russo's study, designed to address the limitation, is the addition of a self-report component in which teachers describe their decision strategies.

METHOD

Subjects

Forty student teachers and 27 practicing teachers participated in the study. The student teachers were enrolled in an elementary education program in the College of Education of a large Southern university. The teachers were elementary school teachers with whom the student teachers worked during their student teaching experiences.

Treatment Variables: Descriptions of Students

Descriptions of 32 hypothetical students were constructed by systematically varying five student characteristics which typically are available to teachers early in the school year when they are making grouping decisions: reading score on a standardized achievement test, self-correcting behavior during oral reading, class participation, social competence, and classroom behavior. Each cue had two levels, so that 25 or 32 student descriptions could be created by forming all possible combinations of the five informational variables. One of the descriptions is shown below:

> Vickie often disrupts class activities with inappropriate behavior. She rarely shares her ideas in lessons and class discussions. She generally gets along well with her fellow students. Vickie recently received a reading score several months below grade level on a standardized achievement test. She corrected substitutions which changed meaning when reading aloud from a second grade reader.

Individual Difference Variables: Conceptions of Reading

Participants' conceptions of reading were assessed using the *Propositional Inventory* (Duffy and Metheny, 1979). This inventory is a 45-item Likert scale which assesses teachers' conceptions along five dimensions which reflect five major conceptual views of reading: linear skills, basal text, natural language, interest, and integrated curriculum.

Dependent Variables: Reading Groups

Each teacher was asked to place the 32 hypothetical students into three reading groups of approximately equal size, using the information available in the description. No instructions were given as to the criteria to use in forming the reading groups.

Procedures

Student teachers participated in the study during a session of one of their courses on the teaching of reading. Teachers participated in small groups at their schools, during the school day. All experimental sessions were held during Spring 1981. Participants' first task was to complete the conceptions of reading inventory. They then were given a brief description of the context in which the grouping decision was to be made. They were asked to assume that it was November and that they had agreed to take over a classroom for a second grade teacher who had taken a leave of absence. Before departing, the teacher had prepared a description of each student in the class, to help in planning the instructional program. Their first planning decision was to place the children in three reading groups. After completing this task, participants filled out a questionnaire in which they were asked to describe the reading groups they had formed, explain the strategies they had used in forming these groups, and discuss any students for whom placement had been a problem.

RESULTS

Conceptions of Reading

Correlations among the subscales of the conceptions of reading inventory clearly indicated that participants were responding similarly to the basal text and linear skills items, and also to the interest-based, natural language, and integrated curriculum items. Duffy and Metheny (1979) reported a similar pattern in their teachers' responses, based on factor analyses of the five subscales and of the individual scale items. They concluded that while the literature suggests that standard reading methods texts can be categorized according to five belief systems, teachers seem to group themselves into two broader categories—a "content-centered" conception (which includes both the basal text and the linear skills conceptions) and a "pupil-centered" conception (which includes interest-based, natural language, and integrated curriculum conceptions). A factor analysis of participants' responses in the present investigation further supported this pattern. Therefore the decision was made to conduct further analyses using the content-centered conceptions rather than the five original subscales.

A multivariate ANOVA was performed to determine whether teachers and student teachers differed in their conceptions of reading. Significant results for both the [MANOVA, Wilks Lambda $= .64, F(2,57) = 15.94$, $p < .001$ and univariate ANOVAS, content: $F(1,58) = 28.42$, $p < .001$; pupil: $F(1,58) = 13.29$, $p < .001$] indicate that there was, indeed, a difference between teachers' and student teachers' conceptions. Student teachers have stronger pupil-centered conceptions than do teachers, while teachers have stronger content-centered conceptions than do student teachers (see Table 1).

Formation of Reading Groups

The written descriptions of reading groups and grouping strategies indicated that virtually all participants had formed groups which they categorized as high, middle, and low. Thus, an Ordinary Least Squares (OLS) analysis could be used to identify participants' decision policies.

The first question asked in the analyses of grouping decisions was whether a single regression equation obtained by pooling data from all teachers and student teachers, accurately represented participants' decision strategies. To answer

TABLE 1
Means, Standard Deviations, and Reliablity Coefficients for Pupil-Centered and Content-Centered Conceptions of Reading (Total Sample, Teachers, and Student Teachers)

	Internal Consistency	Total[1] $\bar{X}$	s	Teachers $\bar{X}$	s	Student Teachers $\bar{X}$	s
Pupil	.87	60.35	11.29	65.59	12.34	56.63	8.90
Content	.90	50.31	11.08	42.84	9.60	55.69	8.78

[1] Total points possible for the pupil-centered subscale is 135; for the content-centered subscale, total points = 90. A low score on a subscale indicates a strong commitment to the conception of reading measured by that subscale.

this question, slopes of all 41 individual regression equations were compared. Since the assumption of equality of regression slopes was violated, $F(330,1742)=3.49$, $p<.01$, a separate regression equation was computed for each participant in order to describe their decision strategy. An examination of these regression equations showed that almost all participants used reading achievement scores (64 of 67) and self-correction of errors during oral reading (47 of 57) in forming their reading groups. Twenty-nine of the participants used information about class participation, 23 considered students' behavior in class, and 10 took into account the students' social competence. In the case of all five variables, students described in positive terms were generally placed in higher reading groups than students described in negative terms.

The two policies most frequently used in forming the reading group were: 1) to consider the reading achievement score and self correction of errors ($N=18$) and 2) to take into account class participation in addition to the two reading cues ($N=17$). For three participants all five pieces of information had a significant effect on the placement of students. At the other extreme, for one teacher and one student teacher, none of the student cues had a statistically significant effect on the grouping decision.

The next question asked of the data was whether professional experience (teacher, student teacher) or conceptions of reading could be used to explain variations among participants' judgment policies. To answer this question, a canonical correlation was computed using the OLS regression weights for the five student characteristics as dependent variables and professional experience, pupil-centered and content-centered conceptions of reading as independent variables. Since the canonical correlation was significant, Wilks Lambda $=.52$, $F(15,144)=2.56$, $p<.002$, regression equations for the five student cues were examined to identify the specific cues for which participants' characteristics explained differences in usage. This examination indicated that participant characteristics played a significant role in determining use of information about class participation [$F(3,56)=8.77$, $p<.05$, and achievement test scores $F(3,56)=2.86$, $p<.05$].

The cue most affected by participant characteristics was class participation, for which 32% of the variation in usage was explained by the regression model. The factor which played a significant role in determining use of class participation information was professional status ($\beta=-.645$, $t=-4.69$, $p<.001$). As the individual regression equations reveal, 63% of the teachers took class participation into account, while only 30% of the student teachers did so.

Participant characteristics accounted for 13% of the variance in the use of both reading achievement test scores and social competence cues. The use of reading achievement scores was significantly affected by professional status ($\beta = .320$, $t = 2.06$, $p < .05$) and content-centered conceptions ($\beta = .445$, $t = 2.77$, $p < .05$). Participants with stronger content-oriented conceptions were more likely to take reading scores into account than were those with weaker content-oriented conceptions. Also, student teachers were more inclined to consider reading scores than were teachers (98% as opposed to 93%), and placed more importance on these scores (as evidenced by larger beta-weights in the individual regression equations). In contrast to the reading achievement score, use of the social competence cue was not significantly affected by any of the three participant characteristics.

Further insight into differences in decision strategies was provided by participants' written descriptions of their reading groups and strategies for forming groups. For example, the finding that no student cues had a statistically significant effect on decision policies of two participants who did not utilize the information provided in forming their reading groups. However, an inspection of their stated policies reveals that these two participants consciously attempted to form heterogeneous reading groups (a finding not apparent from the regression equations, which cannot differentiate between a policy in which cues are ignored and one in which a systematic attempt is made to form heterogeneous groups).

As another example, the written descriptions suggest the thought processes of many participants who used three pieces of information in forming their groups. These teachers and student teachers first divided students into two groups using the characteristic they selected as a primary criterion (e.g., reading achievement scores). These groups were further divided using a second criterion (e.g., self-correction during oral reading). Students scoring high on both variables were placed in the high group; those scoring low on both were placed in the low group. Then, to equalize group size, the additional cue (e.g., class participation) was used to compensate for values along the second dimension. In other words, two or three students with high reading achievement scores, poor oral reading performance, and appropriate class participation were placed in the high group. Similarly, two or three with low reading achievement scores, good oral reading performance, and little class participation were placed in the low group. The remaining students were placed in the middle group.

Self reports for most of the participants whose decision policies contained only the two reading cues indicated that these participants typically utilized decision strategies similar to the one described above. The major difference was in their approach to equalizing group size. Most of the participants reported that they used a combination of two or three of the nonacademic cues, with the result that none of the cues had a statistically significant effect.

DISCUSSION

This study represents an important step toward understanding teachers' conceptions of the reading process and their decisions in the realm of reading instruction. Findings confirm previous investigations, indicating that teachers' conceptions of reading can be grouped into two broad categories, a content-centered conception, and a pupil-centered conception (Duffy and Metheny, 1979). The differences between teachers' conceptions (more content-centered) and student teachers' conceptions (more pupil-centered) in this study can perhaps be explained by the experiences that most probably have shaped these belief systems. Student teachers' conceptions of reading most probably are based primarily on textbook learning and information encountered in classes on the teaching of reading rather than on their student teaching experience. They had been exposed, in classes, to many different approaches to reading instruction. This representative exposure, combined with the predispositions of beginning teachers to be child-oriented (Lortie, 1975) apparently convinced student teachers that the pupil-centered conceptions are the most compelling. In contrast, the district in which participating teachers are employed places a very heavy emphasis on a basal reading program. Teachers are required to use specific basal text and are held accountable for student progress through these texts. It is not surprising, therefore, that teachers participating in this study were found to have strong content-centered (basal text and linear skills) conceptions of reading.

Teachers' strategies for grouping children for reading instruction could be viewed as more complex than those of student teachers. That is, teachers tended to consider the social context and nature of group interactions (i.e., class participation, social competence) to a greater extent than did student teachers, while student teachers placed more emphasis on reading achievement scores. Again, a feasible explanation for this finding centers on differences in experience. Because the teachers have worked with reading groups for much longer than have student teachers, they are more aware of the impact of group dynamics on the learning process. They therefore consider social and behavioral factors in addition to academic factors when forming reading groups.

Another important contribution of this study to our understanding of reading instruction derives from the combination of self-report and policy capturing methodologies. Specifically, this study demonstrated that self-report techniques can enhance our understanding of teacher decision-making by suggesting specific cognitive processes teachers use in making their decisions. The combination of methodologies leads to the conclusion that teachers do, in fact, utilize student characteristics when forming reading groups, and that at least some teachers consider these cues in a sequential, compensatory fashion.

The self-report techniques also suggest ways to improve the design of laboratory experimental investigations. For example, participants in the study reported that the requirement to form exactly three groups of equal size was a constraint uncharacteristic of actual classroom decision making. Future studies, therefore, should allow teachers to determine the number of reading groups they form and the size of each group.

REFERENCES

Anderson, N. H., & Shanteau, J. Weak inferences with linear models. *Psychological Bulletin*, 1977, *84*, 1155–1170.

Bawden, R., Buike, S., & Duffy, G. Teacher conceptions of reading and their influence on instruction. (Research Series No. 47.) East Lansing: Institute for Research on Teaching, Michigan State University, 1979.

Borko, H., & Cadwell, J. Individual differences in teachers' decision strategies: An investigation of classroom organization and management decisions. *Journal of Educational Psychology*, 1982, *74*, 598–610.

Clements, B. S., & Evertson, C. M. Orchestrating small group instruction in elementary school classroom. Paper presented at the annual meeting of American Educational Research Association, New York, 1982.

Duffy, R. E., et al. Measuring teachers' beliefs. (Research Series No. 41.) East Lansing: Institute for Research on Teaching, Michigan State University, April 1979.

Floden, R. E., et al. Responses to curriculum pressures: A policy-capturing study of teacher decisions about content. *Journal of Educational Psychology*, 1981, *73*, 129–141.

Lortie, D. *School Teacher: A Sociological Study*. Chicago: The University of Chicago Press, 1975.

Russo, N. The effects of student characteristics, educational beliefs and instructional task on teachers' preinstructional decisions in reading and math. Unpublished doctoral dissertation, University of California, Los Angeles, 1978.

Shavelson, R. J., and Stern, P. Research on teachers' pedagogical thoughts, judgments, decisions and behavior. *Review of Educational Research*, 1981, *51*, 455–498.

Shulman, L., & Elstein, A. Studies of problem solving, judgment, and decision making: Implications for educational research. In F. N. Kerlinger (Ed.), *Review of Research in Education*. Itasca, IL.: F. E. Peacock Publishers, 1975.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, *28*, 1–39.

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, *6*, 649–744.

Yinger, R., & Clark, C. M. *Selecting instructional activities: policy-capturing and process tracing analyses of teacher judgment*. Paper given at the annual meeting of the American Educational Research Association, New York, 1982.

ROBERT B. RUDDELL
University of California, Berkeley

MARTHA RAPP HAGGARD
Sonoma State University

INFLUENTIAL TEACHERS: CHARACTERISTICS AND CLASSROOM PERFORMANCE

There is great need to understand variables contributing to teaching effectiveness. A substantial body of research indicates that time on task, adjusting difficulty of instruction, appropriate use of feedback, and student praise do influence achievement (Ruddell, 1981). The research in teaching effectiveness, however, has not reached a level of high productivity because of at least three factors. First, past emphasis has largely focused on achievement product rather than achievement process. Second, the learning environment of the classroom has not been given priority attention; and third, little concern has been given to the relationship between teaching style and the learning process.

This investigation was designed to provide information on teaching effectiveness and the instructional process by identifying influential teachers, examining the characteristics of such teachers and analyzing their classroom performance. An influential teacher is defined as a teacher who has been identified by a former student as contributing in a significant manner to the student's academic achievement or adjustment in the school setting between kindergarten and grade eleven. The five questions which served to guide the study follow:

1. What characteristics do influential teachers believe important to good teaching?
2. What characteristics do influential teachers believe to be important in being identified as an influential teacher?
3. What characteristics do influential teachers use in describing teachers who were significant in their own academic preparation?
4. What similarity or difference exists in the way students and their influential teachers prioritize characteristics of good teaching?
5. What profile best describes the influential teachers studied; what is the nature of the personal interaction patterns in the classrooms of influential teachers, and are influential teachers effective teachers?

METHOD

This longitudinal study was conducted in a Northern California school district. The Ss were part of a sample of 522 primary grade students who had participated in a study in the 1970–71 academic year (Ruddell and Williams, 1972). A detailed examination of records revealed that 132 of these students remained in the district nine years later. In a previous study (Ruddell, 1980) data were collected to identify the way in which these students used literacy skills both in and out of school. In addition students were asked to identify teachers who had greatly influenced them during their academic careers and to specify the nature of this influence. The student responses were analyzed and five influential teacher descriptors were formulated. These were as follows:

1. Personal Characteristics—behavior which demonstrated sensitivity, openness, supportive attitude, and sense of humor;
2. Understands Learner Potential—behavior which reflects understanding and uniqueness of each student;
3. Attitude Toward Subject—behavior which demonstrates enthusiasm for subject and knowledge of subject matter;
4. Life Adjustment—behavior showing concern and sensitivity to student problems both academic and personal;
5. Quality of Instruction—behavior demonstrating expertness in subject, clarity of expectations, and ability to present subject matter effectively.

High and low achieving students were defined on the basis of those Ss in the top 30% and bottom 30% of scores on a standardized achievement test.

The project staff was successful in locating 79 teachers who still remained in the school district. Of this number, 15 elementary and 25 secondary teachers were willing to cooperate in the research study.

A detailed questionnaire and interview schedule was designed to study the characteristics of the influential teachers. The 24 item questionnaire and the 10 item interview were piloted, revised, and administered to the 40 influential teachers.

Classroom performance was evaluated using videotaped lessons presented by four influential teachers named in the original study. All four teachers had been named by more than one student, and total student comments included all five categories of teacher characteristics. In the videotapes, each teacher presented the book, *Alexander and the Wind-Up Mouse* (Lionni, 1968) to six children ramdomly chosen from their own primary grade classroom. They had been asked to use the book as they would normally use it with their entire class.

Four elements of personal interaction patterns were assessed using a seven-point scale. The four elements are

Classroom Communication System (verbal and nonverbal)—ratings range from (1) open, two-way interaction; seeks clarification and resolution, to (7) closed interaction, defensive, hostile.

View of Self in Relationship to Others—ratings range from (1) equal, secure; strong sense of self and classroom purpose, to (7) inferior, weak self definition, rejected.

Management Style/Motivation—ratings range from (1) cooperatively determined action/high motivation, to (7) use of authority, reprimand/very low motivation.

Problem Solution Approach/Attitude Toward Learning—ratings range from (1) confidence in self, open consultation, intellectually curious, internal locus of control/positive, to (7) overwhelmed by problems, little curiosity, external locus of control/negative.

Teaching effectiveness was evaluated by comparing teacher behavior to the instructional functions of effective teaching described by Fisher *et al*. (1980). The four instructional functions assessed were 1) *Presentation*—includes planned explanation, unplanned explanation, and providing directions; 2) *Monitoring*—includes observing student activities and questioning students; 3) *Feedback*—includes feedback about student academic response and feedback for the purpose of controlling student attention to the task; and 4) *Context*—

general focus or orientation of the teacher. Context may be academic, affective, or a combination of the two. Viewing and evaluation were done independently by two raters. Evaluations were compared and differences resolved through discussion. Interrater reliability was at the 95% level.

RESULTS

The first research question examined characteristics influential teachers believed important to good teaching. Teacher responses were categorized using the five descriptors and were found to order in the following manner: Personal Characteristics, 47%; Understands Learner Potential, 18%; Attitude Toward Subject, 18%; Quality of Instruction, 16%; and Life Adjustment, 1%. Clearly, Personal Characteristics was the most valued quality in the influential teachers' perception of good teaching with Quality of Instruction and Life Adjustment rating comparatively low.

The second question explored characteristics which the teachers believed were critical in their identification by students as influential teachers. These results closely parallel those findings for the first question ordered as Personal Characteristics, 48%, Understands Learner Potential, 25%, Quality of Instruction, 11%, Attitude Toward Subject, 10%, and Life Adjustment, 6%. These results were not surprising; again, however, the great differential between the high Personal Characteristics ranking and the low Quality of Instruction ranking was an unexpected finding.

In regard to the third question, the influential teachers identified the characteristics of their own significant teachers as follows: Personal Characteristics, 44%; Understands Learner Potential, 27%; Attitude Toward Subject, 23%; Life Adjustment, 5%; and Quality of Instruction, 1%. When these ratings are compared with the teacher responses to "Why I was named as an influential teacher," a high degree of similarity is present for each descriptor.

The fourth question examined the similarities and differences between the influential teacher ratings completed by the high and low achieving students. The ratings between high and low achievers were found to be remarkably similar as shown for the following descriptors, respectively; Personal Characteristics, 14% and 13%; Understands Learner Potential, 7% and 4%; Attitude Toward Subject, 15% and 13%; Life Adjustment, 19% and 23%; and Quality of Instruction, 45% and 47%.

The surprising finding in these data is in the contrast between student ratings and the influential teacher ratings. While the high and low achieving students identified Personal Characteristics at 14% and 13%, respectively, the influential teachers provide a 47% response for this factor. Conversely while the influential teachers rate Quality of Instruction at only a 16% frequency level, the students view this factor as an important one, with responses at the 45% and 47% levels. A substantial difference was also noted for the Life Adjustment descriptor with teachers responding at a 1% level and high and low students at the 19% and 23% level, respectively. Much greater similarity is noted for the descriptors Attitude Toward Subject, 18% vs. 15% and 13%, while greater variation is present for Understands Learner Potential, 18% vs. 7% and 4%, in comparing influential teachers with high and low achievers, respectively.

These findings suggest that the influential teachers place a much higher priority on Personal Characteristics in their perceptions of factors important to good teaching, in their view of self, and in their own significant teachers than do high and low achieving students. In contrast, the students place a much higher priority on Quality of Instruction and Life Adjustment than do their influential teachers.

The fifth question focused on formulation of the profile which would best describe the influential teachers in the study. Open communications systems were observed for all four teachers, characterized by attentive response to the children, high levels of eye contact, and openness in facial expression and body posture. Differences in ratings (1 for teacher A, 3 for Teachers B and C, 4 for Teacher D) generally grew from the way the teachers directed discussion of the book, and how they worked toward resolution of questions and issues. Teachers A, B and C all fell within the general range of "equal and secure," with view of self ratings of 2, 1, and 2, respectively. They appeared confident in themselves as teachers and secure in their relationships with the children, even in the unusual circumstance of presenting a lesson for videotaping. Teacher D (rated 4) appeared much less secure in his role. Differences in management styles were apparent throughout the teachers' interactions with the groups. Teacher A, who received a rating of 1, developed a discussion which became a group conversation; Teachers B and C, rated 3, clearly directed the discussion through use of praise and admonition. Teacher D, rated 4, used no direct praise, but did admonish when extraneous talking and play made it difficult for individual comments to be heard. Intellectual curiosity and locus of control were expressed in teacher response during the discussions. Respective ratings for Teachers A, B, C, and D were 1, 2, 3, 4.

Profiles of teaching effectiveness were based upon analysis of the four areas of presentation, monitoring, feedback and context. With regard to presentation, all of the teachers had some degree of planned explanation and discussion in their lessons. Teachers B and C demonstrated the clearest structure, and Teacher D the least. Teachers A, B, and C gave the most explicit directions for group procedures. Monitering activities were concentrated mainly on questioning students about both content and feelings. Teacher A's first question after reading was open-ended, requiring either spontaneous or probe-response explanation. Teacher B's initial question was content-specific, followed by opinion questions. Teacher C asked the greatest number of questions, most of which were designed to determine literal recall. Teacher D concentrated all of the discussion on feelings, with little of no attention to content. All of the teachers provided academic task feedback, either directly or indirectly. Teacher A's emphasis on extended response provided implicit feedback. Teachers B and C were the most direct, with explicit positive responses or continued search for answers. Teacher D's responses were indirect and implicit, but were commensurate with the type of questions asked. Context analysis indicated that Teachers B and C appeared to be primarily academically oriented. Their lessons had clear academic goals and discussion was structured toward those goals. Teacher D's lesson was focused toward affect, and instructional goals were not as readily apparent. Teacher A demonstrated a combined emphasis on academic content and affect, with academics having first priority.

CONCLUSION

Based on the findings of the study the three conclusions are warranted. First, the influential teachers place Personal Characteristics, Understands Learner Potential and Attitude Toward Subject in highest priority. The descriptors of Quality of Instruction and Life Adjustment rank much lower in their priority system. There is a very close relationship between the influential teachers' perception of qualities important to good teaching and their perception of self, as well as those qualities identified in their own significant teachers. One may hypothesize that a strong modeling effect exists in the influential teacher which may directly affect their students, and in turn future teachers.

Second, there is a marked difference in the way students prioritize good teaching characteristics when contrasted with their influential teachers. Both high and low achieving students view Quality of Instruction and Life Adjustment as most important to good teaching, characteristics which were ranked lowest by their influential teachers. Conversely Personal Characteristics and Understands Learner Potential were ranked lowest by both types of students but highest by the influential teachers.

Third, in constructing a profile of the influential teacher, based upon this representative sample, it appears that influential teachers, although exhibiting quite different personal interaction styles, are more alike in their ability to meet student needs than surface differences might indicate. While management styles and teaching approaches varied from highly structured, formal group discussion of content to highly unstructured, more informal exploration of affect, it is important to remember that *all* of these teachers were identified as influential by their students; it may be that such variation is necessary from classroom to classroom if the needs of widely different children are to be met. These findings also suggest that not all influential teachers are effective teachers, at least as currently defined, and that the impact they have on students' lives must be measured by variables other than the traditional ones of academic success and student achievement.

This study has identified critical instructional characteristics both from the perspective of influential teachers and high and low achieving students, and has analyzed two elements of classroom performance: personal interaction patterns and teaching effectiveness. These variables deserve careful exploration in future research. Future research efforts must also place greater emphasis on the study of the teacher and the student as they participate in the instructional process if our understanding of learning and effective instructional change is to be enhanced.

REFERENCES

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. *Teaching behaviors, academic learning time, and student achievement: an overview*. In C. Denham & A. Lieberman (Eds.), *Time to learn*. U.S. Department of Education, National Institute of Education, 1980, 7–32.

Lionni, L. *Alexander and the wind-up mouse*. New York: Pantheon Press, 1968.

Ruddell, R. *Literacy achievement profiles and literacy use—high and low achievers*. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives on reading research and instruction*. 29th Yearbook of the National Reading Conference, 1980, 292–300.

Ruddell, R. *Teacher impact on student achievement: implications for reading instruction*. Paper presented at the International Reading Association, New Orleans, 1981.

Ruddell, R., & Williams, A. *A research investigation of a literacy teaching model: project delta*. Washington, D.C. U.S. Department of Health, Education, and Welfare, Office of Education, Project No. 005262, 1972.

WILLIAM H. RUPLEY
TIMOTHY R. BLAIR
BETH S. WISE
Texas A & M University

SPECIFICATION OF PROMISING TEACHER EFFECTIVENESS VARIABLES FOR READING INSTRUCTION

It is logical to assume that research results about effective teaching of reading could result in improving the quality of classroom reading instruction. The accumulated knowledge about effective teaching of reading is at a point where it can begin to serve as a basis for describing many of the instructional process variables associated with pupils' reading achievement (Anderson, Everston, and Brophy, 1979; Heilman, Blair, and Rupley, 1981; McDonald and Elias, 1976; Rosenshine, 1976; Rupley and Blair, 1981). Recent research efforts (Anderson, Everston, and Brophy, 1979; Stallings, Needels, and Stayrook, 1979) have demonstrated that teachers trained to use teaching strategies associated with effective reading teachers obtain greater pupil achievement gains in reading than do untrained teachers. It is also apparent in the early reading research literature that much of what is claimed as new and insightful today were primary areas of investigation during the turn of the century. For example, features of recent effective reading instruction inquiry such as allocated time for instruction, opportunity to learn, time on task and content covered were also investigated in the 1920s and 1930s (Ford, 1924; Gates, 1936; Gray, 1933; Reinoehl, 1922; Woody, 1923). Many of the recommendations based on the findings of this early research parallel closely those that are offered currently as means for improving the quality of reading instruction. Why is it, then, that most teachers of reading are not characterized as relying predominantly on the use of such effective strategies in their reading instruction?

There are several explanations for the lack of implementation of effective reading instruction inquiry results. One credible explanation is the limited attention given to whether or not classroom teachers perceive a need for changing their reading curriculum and instruction. If the cumulative results of teacher effectiveness research are to significantly impact reading instruction, increased attention needs to be given to describing existing reading instruction, specifying what needs to be modified, and ascertaining how best to communicate promising research findings to classroom teachers.

The major purpose of the present study was to describe selected reading instructional strategies and procedures used by elementary school teachers. Primary focus of the study was on reported teaching behaviors and procedures used in the teachers' reading instructional programs. These data could then be compared with recommendations found in the reading teacher effectiveness literature. A secondary purpose was to compare identified instructional strategies and procedures used by the teachers in relation to their level of success status over a six year period.

METHOD

Subjects

Over a period of six years, at two year intervals, all third and sixth grade teachers who taught reading in self-contained classrooms in a large midwestern school district were identified as high, average, or low effective teachers of reading. Teacher's success status was determined by comparing pupils' end of year reading achievement (Total Reading on the SRA Achievement Test) with their predicted end of year reading achievement. Least squares prediction procedures, utilizing the standard error of estimate, were used to identify teachers success status stability for each of the two year intervals. Teachers were then categorized by their stability of effectiveness. That is, high effective teachers were those who for each of the two year intervals produced greater than predicted pupil achievement. Identical procedures were used to identify teachers for average and less effective success categories.

Eleven third grade teachers were randomly selected from each of the three success status stability categories for a total sample of 33 teachers. Due to attrition, all sixth grade teachers were included in the sample, resulting in eight high, fourteen average, and nine low effective teachers.

Data Collection

Teachers in the sample were administered a questionnaire designed to gather information about their classroom reading instruction. Major categories in the questionnaire were skill sequencing, record keeping, group organization, group instruction, and material usage. Teachers responded to questionnaire items by either selecting a description, writing an explanation, or identifying a frequency for each major category. Ninety-six percent of the questionnaires distributed were completed and returned.

Analysis

Data were summarized quantitatively by teachers' success status and grade level. Frequencies for each questionnaire category were calculated and mean frequencies computed for each success status by grade level. This resulted in data summaries that were compared by success status categories at both third and sixth grades. Mean frequencies were also computed for each questionnaire area of interest for all third and all sixth grade teachers in the sample, which allowed for a description of reading programs at each grade level.

Specification of frequencies for instructional procedures by success status at each grade level resulted in comparative data at three levels of teacher effectiveness. Descriptive data summaries at each grade level provided a point of comparison for success status levels within and between grade levels.

RESULTS

Success status differences for teachers' use of basal series were noted at both grade levels. Seventy-three percent of the high success third grade teachers reported the use of a basal reader series; forty-five percent of both average and low success status third-grade teachers reported using a basal series for reading instruction. At sixth grade a similar pattern was noted; seventy-five percent of high, sixty-four percent of average, and forty-four percent of low success status teachers reported basal series usage for their reading instruction.

The average number of reading groups by grade level and success status revealed some notable differences. Within grade levels, high success teachers reported fewer numbers of reading groups than did low success teachers. At third grade, high success teachers had an average of 2.4 reading groups and low success teachers averaged 5.0 reading groups. High success sixth grade teachers had an average of 2.1 reading groups compared to 3.3 reading groups reported for low success status sixth grade teachers.

Two categories of the questionnaire dealt with planning instruction and diagnosing pupils' reading: use of skill sequence charts in instructional planning, and frequency of record keeping and use of teacher-prepared and basal level tests.

Sixty-four percent of both the high and average success third grade teachers reported the use of reading skill sequence charts in their instructional planning. Low success status third grade teachers indicated a 36 percent usage of sequence charts. Types of charts used by these third grade teachers were commercial (basal) and teacher prepared. All of the high success status teachers using sequence charts reported using only a commercial sequence chart.

Sixth grade teachers reported less use of sequence skill charts than did third grade teachers, 29 percent compared to 55 percent. Less than 20 percent of high and low success status sixth grade teachers reported the use of skill sequence charts. Fifty percent of average success teachers used skill sequence charts.

Frequency of record keeping on pupils' reading and use of teacher-prepared tests and basal level tests varied little between third and sixth grade teachers. At both third and sixth grades most teachers reported keeping weekly records. In comparing teachers by success status, the greatest difference for third grade was between high and average success status teachers, who reported 67 percent and 36 percent, respectively, for weekly record keeping. At the sixth grade level, 63 percent of high success status teachers reported keeping weekly records compared to 71 percent of average and 44 percent of low success status teachers. The greatest difference at sixth grade was for monthly record keeping—44 percent of low success, 15 percent of average success, and 12 percent of high success status teachers were placed in this category.

Teachers' grouping of pupils for reading instruction was another major section of the questionnaire. Teachers' responses to items dealing with group flexibility, criteria used for changing group structure, and group skill instruction revealed some interesting differences. All of the third grade teachers reported using flexible grouping; however, 71 percent of the sixth grade teachers utilized flexible grouping. The greatest difference in the use of flexible grouping at the sixth grade was for average effective teachers who noted a 50 percent usage, compared to 100 percent of high success and 77 percent for low success teachers. At third grade, 55 percent of the high effective teachers reported using pupil progress as the major criterion for placing a pupil in a different group contrasted with 27 percent of the low success teachers. Sixty-three percent of high success sixth grade teachers used pupil progress as a means for changing pupils' group placement, contrasted with 29 percent of low success teachers.

In addition to group flexibility, data on teacher's use of skill grouping and the reading instruction of such groups were gathered. Success status differences for teachers who did use skill grouping were as follows: high third grade, 82 percent, average third grade, 45 percent, low third grade, 64 percent. Frequency differences for use of skill grouping by success status were also noted at sixth grade: 75 percent high, 43 percent average, and 44 percent low.

The focus of instruction for these skill groups varied between grade levels. Phonics, vocabulary, and comprehension were reported receiving emphasis by 36 percent or more of third grade teachers using skill grouping. Comprehension and vocabulary were emphasized by 36 percent of sixth grade teachers in their skill group instruction. Comprehension and vocabulary were also the instructional areas that varied the most within grade levels by success status categories. Of the high success third grade teachers who used skill grouping, 66 percent used it for vocabulary instruction and 22 percent for comprehension instruction. One hundred percent of the high effective sixth grade teachers reported the teaching of comprehension and 83 percent reported the teaching of vocabulary in skill groups. At sixth grade for both average and low success status teachers, 50 percent of the teachers who grouped for skill instruction focused on vocabulary and comprehension.

DISCUSSION

The data support many of the current research findings in teacher effectiveness and provide new insights into effective instructional procedures. Findings of the present study parallel closely recent teacher effectiveness research. Effective teachers at both third and sixth grade levels used fewer reading groups than did less effective teachers. This finding is consistent with other investigations and time on task studies (Rosenshine, 1979). Opportunity to learn (Guthrie, Martuza, Seifert, 1976) was reflected in teachers' reported use of sequence skill charts, which provided instructional development related to valued reading outcomes. More effective teachers relied upon commercial skill charts connected directly with the basal reader series, while less effective teachers relied more heavily on teacher prepared sequence charts.

Another major variable associated with quality reading instruction is use of diagnosis (Heilman, Blair, Rupley, 1981). All teachers in the study kept records on pupils' reading, but more effective teachers reported more emphasis on basal reader diagnostic level tests and less effective teachers concentrated their efforts on teacher-made diagnostic instruments. Related to diagnostic record keeping, all teachers used flexible grouping procedures. However, a major discrepancy was found between effective and less effective teachers. Effective teachers more often used the criterion of progress to change pupils from one group to another. In constrast, less effective teachers used the criterion of lack of progress to change group membership.

The results regarding the use of the basal reader are related to the data reported for record keeping, sequence charts, and diagnostic tests. Effective teachers reported to have used basal readers more than did less effective teachers. This finding can be related to the importance of teacher-directed instruction, where identifying instructional objectives, pacing instruction, monitoring pupils' performance, and maintaining pupils' engagement in the lessons are more likely to be efficiently accomplished with a systematic program of reading instruction. While some criticism has been cited in the literature on reliance upon the basal reader (Durkin, 1978–79), a crucial question seems to be, how does one use the basal effectively? Additional inquiry into how effective teachers use the basal to the enhancement of learning needs to be conducted.

IMPLICATIONS

The results depicting specific reading instructional practices according to success status are important because they provide descriptive data on the teaching of reading. As stated previously, many of the current research results in teacher effectiveness have been studied over the past fifty years. In searching for reasons why such information has not had the impact it should have in our schools, the authors feel that little attention has been paid to teacher's perceptions of teacher effectiveness.

Teachers will not change their instruction in light of research findings if they perceive their own teaching as being congruent with pupils' success. Indeed, if teachers are to seriously consider new techniques or strategies, they must be shown actual data representing their instruction in comparison to student achievement gain. Thus, the data in this study could be presented to teachers as descriptive in nature and compared with current teacher effectiveness research results. This is a necessary step in the change process. Teachers must be shown that a gap exists between what actually is happening in classrooms and what research shows is productive for student learning.

REFERENCES

Anderson, L., Everston, C., & Brophy J. An experimental study of effective teaching in first grade reading groups. *Elementary School Journal*, 1979, *79*, 193–223.

Durkin, D. What classroom observations about reading compehension instruction. *Reading Research Quarterly*, 1978–79, *XIV (4)*, 481–533.

Ford, F. A. Diagnostic supervision. *Journal of the Louisiana Teachers Association, I*, April 1924, 30–35.

Gates, A. I. The necessary age for beginning reading. *Elementary School Journal*, 1937, *37*, 497–508.

Gray, W. S. *Improving instruction for reading*. Chicago, IL: University of Chicago Press, 1973.

Guthrie, J. T., Martuza, V., & Seifert, M. *Impacts of instructional times in reading*. Newark, DL: International Reading Association, 1976.

Heilman, A., Blair, T. R., & Rupley, W. H. *Principles and practices of teaching reading*, (5th ed.). Columbus, OH: Charles E. Merrill Publishing Co., 1981.

McDonald, F., & Elias, P. *The effects of teaching performance on pupil learning*. Beginning Teacher Evaluation Study: Phase II, 1973–74. Final Report: Vol. 1. Princeton, NJ: Educational Testing Service, 1976.

Reinoehl, C. M. Analytic survey of state courses of study for rural elementary schools. Washington, D.C.: *Bureau of Education Bulletin No. 42*, 1922.

Rosenshine, B. V. Recent research on teaching behaviors and student achievement. *Journal of Teacher Education*, 1976, *27*, 61–64.

Rosenshine, B. V. Content time and direct instruction. *Research on Teaching: Concepts, Findings and Implications* (H. Walberg and P. Peterson, eds.). Berkeley, CA: McCutchan Publishing Co., 1979.

Rupley, W. H., & Blair, T. R. Specification of reading instructional practices associated with pupil achievement gains. *Educational and Psychological Research*, in press.

Stallings, J., Needels, M., & Stayrook, N. *The teaching of basic reading skills in secondary schools. Phase I and phase III*. SRI International, Menlo Park, CA, 1979.

Woody, C. Practices in Teaching First-Grade Reading in the Public Schools of Michigan. *Bureau of educational reference and research bulletin 58*. Ann Arbor, MI: School of Education, University of Michigan, 1923, 34.

PATRICK SHANNON
University of Houston

TEACHERS' SELF-PERCEPTIONS AND REIFICATION OF INSTRUCTION WITHIN READING INSTRUCTION

Perhaps the finding upon which reading researchers most often agree is the one suggesting that elementary-grade teachers rely on commercial materials during their reading instruction (Duffy, 1982). From Austin and Morrison in 1963 to EPIE in 1977, the results have been uniform; eighty-five to nintey percent of elementary teachers use commercial materials during most of their reading instruction. If these observations are correct, the question of why teachers rely so heavily on commercial reading materials becomes important in attempts to understand the reading instruction most American students receive. This study was an attempt to discover rationales for why teachers use commercial materials so often and with apparent confidence.

A three step approach was used toward that end. First, a disclosure model of reading programs was developed to explain teachers' behavior. Second, a study of a reading program was conducted in one school district. Third, the usefulness of the disclosure model was judged according to the data gathered in the empirical investigation. (See Shannon, 1981, for further discussion of each of the steps.)

Step One: A Disclosure Model of Reading Programs

A disclosure model, according to Ramsey (1964), is an explanation of a phenomenon from a perspective not usually taken. That is, a different perspective can lead to insights about the phenomenon. A disclosure model of reading programs will not predict the exact workings of all programs, rather it will explain basic characteristics common to each. Georg Lukács' Theory of Reification (1971) provided a theoretical basis for such a disclosure model. Lukács sought to explain "institutional behavior" as a dialectic among three factors: formal rationality, reification, and alienation. Formal rationality is a tendency to use abstract reason as the sole criterion to organize and to implement an activity designed to lead to a previously designated goal. Reification is a treatment of an abstraction as a concrete thing or immutable procedure. Alienation is a process of separation between people and some other quality assumed to be related in natural circumstances.

The disclosure model of reading programs works from three major tenets. First, American elementary-grade reading programs are organized according to the principles of formal rationality. In other words, they are bureaucracies. Second, these rationalized reading programs are based on the reification of reading instruction as the application of commercial reading materials. Third, the combination of the rationalized reading program and reification of reading instruction forces the objective alienation of classroom teachers from the control of their reading instruction. Based on these three tenets, one can conclude that the teachers' role during reading instruction is limited to the systematic application of commercial materials according to guidebook suggestions. For teachers, the use of commercial materials becomes an end in itself rather than a means to an end.

Step Two: Why Teachers Use Commercial Materials in One School District

The empirical study was divided into two investigations. Subjective factors for teachers use of commercial materials were probed using questionnaires and interview schedules. The survey instruments were designed to test four hypotheses. 1) Teachers are not involved with their reading instruction (Durkin, 1978–79). 2) Teachers believe the commercial materials can teach a student to read without aid or interference from teachers (Austin and Morrison, 1963). 3) Teachers believe that the materials embody scientific truth (Barton and Wilder, 1964). 4) Teachers think they are fulfilling administrative expectations when they use the materials (Chall, 1967).

Objective factors were investigated by a survey of superordinate personnel, with informal observations of interactions between teachers and administrators over a one year period, and through an examination of the school district's printed explanations of their reading program. One objective hypothesis was tested: Administrators expect teachers to use commercial materials according to prescribed procedures.

Development and Distribution of the Survey Instruments

After two pilot studies, twenty forced-choice items were selected for the final version of the teacher's questionnaire. Five items tested each hypothesis. A five point Likert-type format was used to record respondents' acceptance of each hypothesis: a selection of one signified strong disagreement with the hypothesis; three suggested moderate agreement; five meant strong agreement. For example, teachers were asked to respond to this item: The sequence of goals for teaching is scientifically valid within basal reading programs. A response of three was interpreted as moderate agreement with the third hypothesis: commercial materials embody scientific truth.

Additionally, three open-ended questions were included in the questionnaire to allow teachers to elaborate on their perceptions and provide rationales that the investigator had not considered previously. These questions asked teachers to explain "why," "how," and "when" they used commercial materials. Alternate forms of the questionnaire were prepared for reading teachers and administrators.

Interview schedules were developed to check the stability of the information gathered by the questionnaires. One forced-choice item for each hypothesis was selected randomly and included in the interview schedule to meet this concern. The three open-ended questions from the questionnaire were used as the initial portion of the interview to relax the respondent and to probe the depth of their responses to the "why" question.

Appropriate questionnaires were distributed to 539 classroom teachers, 26 reading teachers, and 26 administrators in a large Midwestern school district which served as a national model for the Right to Read Program. Subsequent to the collection of questionnaires, interviews were held with 26 teachers, 3 reading teachers, 3 administrators, and 1 central administrator.

RESULTS

Eighty two percent of the questionnaires were returned. Table 1 shows that each group accepted each hypothesis, at least, moderately. All means were at or above 3.0. A comparison of group means within hypotheses demonstrated sig-

TABLE 1
Mean Scores by Groups for Forced-Choice Items from the Questionnaires

	Hypotheses			
	1*	2	3	4
Teachers (445)	3.00	3.65	3.76	4.21
Reading Teachers (23)	3.23	3.66	3.52	3.57
Administrators (18)	3.17	3.82	4.12	3.60

*1. Teachers are not involved with their reading instruction.
2. Teachers believe the commercial materials can teach.
3. Teachers believe that the materials embody scientific truth.
4. Teachers think they are fulfilling administrative expectations.

nificant differences between means for responses to hypotheses three and four ($F2, 483 = 4.12$, $p < .05$; and $F2, 483 = 18.17$, $p < .001$, respectively). *Post hoc* comparisons (Fisher's Least Significant Difference) suggested that administrators agreed more strongly than reading teachers that commercial materials are based on science ($p < .01$). Classroom teachers agreed more strongly than either administrators or reading teachers that teachers are doing what they are told when they rely on commercial materials ($p < .01$).

A second comparison of means was made to see if a group preferred one hypothesis over another. Classroom teachers preferred administrative expectations as a rationale ($F1, 483 = 228.16$, $p < .001$ using Geiser-Greenhouse Conservative F test). Administrators preferred two rationales: commercial materials can teach reading, and the materials embody scientific truth ($F1, 54 = 14.28$, $p < .01$). Administrators' responses to items testing these hypotheses were correlated significantly ($r = .54$). (These were the only responses that were significantly correlated for any group.) The administrators' mean for hypothesis three, however, was the only mean significantly greater than the mean for hypothesis four. The forced-choice items from the questionnaire were highly correlated with the items from the interviews ($r = .84$).

Three scorers judged the open-ended "why" question to determine the primary hypothesis mentioned (interscorer reliability was $r = .94$ for fifteen percent of the responses). Seventy percent of the teachers stated administrative expectations as the reason they use commercial materials. Reading teachers responses were spread evenly across three hypotheses (hypothesis three was not mentioned). Fifty-six percent of the administrators stated the instructional powers of the materials as the primary reason that teachers use them. The open-ended responses during the interviews were highly similar.

The qualitative results for the investigation of objective factors were divided into seven categories: organization of personnel, goals for instruction, assumptions of instruction, pace of instruction, use of materials, selection of materials, and assessment procedures. Essentially, the reading program was based on a mastery learning model with the goals and procedures for reading instruction abstracted exactly from one set of commercial materials. The personnel were organized hierarchically to deliver that program. Central administrators made most important decisions which principals and reading teachers implemented. Teachers were encouraged to use the materials and to follow the directions in the guidebook "like a Bible." The pace of instruction was described as the only factor under the control of teachers, however, administrators examined biweekly students' scores on criterion referenced tests which the commercial publishers provided. Standardized tests were administered only to sixth grade students who were to graduate from elementary school.

The textbook selection process was the only major point of disagreement between faculty and administrators. Most classroom teachers and reading teachers said that administrators selected the materials, and then, convened a committee to confirm their choice. Administrators said that everyone had an opportunity to participate meaningfully.

Step Three: The Adequacy of the Disclosure Model of Reading Programs

The responses to the questionnaire are fairly easy to interpret. Most teachers thought they were fulfilling the expectations of administrators when they used the commercial materials. Indeed, administrators expected teachers to use commercial materials as suggested in the commercial guidebooks seemingly because they believed the materials can teach reading according to scientifically tested methods. While reading teachers accepted each hypothesis moderately, they did not prefer one hypothesis to another. The difference between reading teachers' responses and those of the other groups might be attributed to a greater number of courses concerning reading instruction. Reading teachers, on average, completed three more courses than either administrators or classroom teachers. This coursework seemed to affect their responses. For example, reading teachers with more coursework were less likely to accept the scientific validity of the commercial materials ($r = -.50$). Classroom teachers and administrators, on the other hand, demonstrated the opposite tendency ($r = .06$ and $.59$, respectively). Educational background did seem to explain the difference between reading teachers' and administrators' responses to items which tested hypothesis three.

These results are consonant with the disclosure model. Administrators sought formal rationality in the reading program through a mastery learning model and strict application of commercial materials. This attempt to actualize formal rationality had a predictable effect on the teachers and administrators; at times, the seem substantially irrational. Administrators treated problems that warranted changes in the rationalized reading program as insolvable. For example, administrators refused to allow third grade teachers to postpone several objectives that proved difficult for even their best students. After three years of argument, teachers were permitted to defer the objectives because the authors of a new edition of the commercial materials had sequenced the objectives as the teachers suggested. Furthermore, the logic that a single set of commercial materials is required for continuity of goals is only internally consistent. A variety of materials would not necessarily compromise the scope and sequence of goals, it would, however, place more control in the hands of the classroom teachers—an anathema to a rationalized program.

The classroom teachers could be considered alienated from their reading instruction in two objective ways. Formal rationality meant that administrators set goals, establish procedures, and influence the pace of instruction. Thus, classroom teachers were separated from control of their reading instruction. Moreover, their development as reading instructors was limited to becoming more proficient in the application of the commercial materials. Evidence for teacher's conscious recognition of this alienation was sparse. However, seventy percent of the teachers agreed strongly that others could teach reading in the same way they do. In this sense, classroom teachers considered themselves alienated.

A closer inspection of classroom teachers' responses suggested that they reify reading instruction also. While classroom teachers were required to use the designated commercial materials, they stated that they would use some commercial materials as the source of their reading instruction even if they were free to choose any means for that instruction. When prodded during the interviews, all twenty-six teachers related that commercial materials supplied the continuity, the instruction, and the assessment methods for their reading instruction. They stopped short of saying the commercial materials can teach.

The disclosure model, which suggests a relationship among reification, rationalization, and alienation, was useful in uncovering the essence of why teachers use commercial reading materials. Yes, teachers were expected to use the materials, but they also sought to use the materials.

REFERENCES

Austin, M., & Morrison, C. *The first r*. New York: Macmillan, 1963.

Barton, A., & Wilder, D. Research and practice in the teaching of reading. In Miles (Ed.), *Innovations in Education*. New York: Columbia University Press, 1964.

Chall, J. *Learning to Read: The Great Debate*. New York: McGraw-Hill, 1967.

Duffy, G. Commentary. *Reading Research Quarterly*, 1982, *17*, 295–300.

Durkin, D. What classroom observation reveals about reading comprehension instruction. *Reading Research Quarterly*, 1978–79, *14*, 481–533.

Educational Products Information Exchange (EPIE). *Report on a national study of the nature and the quality of instructional materials most used by teachers and learners*. Technical Report No. 76, New York: EPIE Institute, 1977.

Lukács, G. *History and Class Consciousness*. Boston: MIT Press, 1971.

Ramsey, I. *Models and Mysteries*. Cambridge University Press, 1964.

Shannon, P. *Teachers' Self-Perceptions and Reification of Instruction within Reading Instruction*. An unpublished doctoral thesis, University of Minnesota, 1981.

JOHN C. STANSELL
R. KAY MOSS
Texas A & M University

CAROL P. ROBECK
TLC The Learning Center, Inc.

THE DEVELOPMENT OF THEORETICAL ORIENTATION TO READING AMONG PRESERVICE TEACHERS: EFFECTS OF A PROFESSIONAL TRAINING PROGRAM

The purpose of this study was to examine the impact of professional training, including campus-based coursework and field experiences, upon the theoretical orientations of students in an undergraduate reading specialization program.

The Concept of Theoretical Orientation

Harste and Burke (1977) have hypothesized that "both the teaching and the learning of reading are theoretically based" (p. 32). That is, both teachers and students behave according to stable and consistent assumptions about the nature of the reading process and its acquisition. Harste and Burke identified three theoretical orientations which they labelled *phonics*, *skills*, and *whole language*. These three differ on the dimension of the size of the language unit seen as critical in reading and learning to read. In a phonics orientation, letter-sound (graphophonemic) relationships are critical in the instructional program. A skills orientation views words as critical units and focuses attention upon word recognition and identification techniques. A whole language orientation entails seeing sentences, paragraphs, and discourses as critical units and focuses attention on intrasentence and intersentence syntax and semantics.

As DeFord (1978) has pointed out, these orientations also differ on another dimension. DeFord described the phonics and skills orientations as arising from a mechanistic theory model (Steiner, 1977) wherein the phenomena to be understood are seen as analogous to machines with each part having a fixed, unalterable function and the whole being the sum of its parts. In such a model, understanding a phenomenon would be possible through separate examination of the several parts. Phonics and skills orientations reflect this model through a pedagogical focus upon separate and somewhat isolated examination of "parts" of written language—letters and words (Burke, 1972; McCracken & Walcutt, 1963; Robinson, Monroe, & Artley, 1962).

On the other hand, a whole language orientation arises from an organismic theory model (Steiner, 1977) wherein phenomena are viewed as analogous to organisms. An understanding of a phenomenon from this perspective involves the notion that parts have alterable functions and are interactive and interdependent in maintaining the function of the whole. A whole language orientation reflects this model through a pedagogical focus upon the interactive and interdependent nature of cue systems in written language (Aaron, Artley, Goodman, Huck, Jenkins, Manning, Monroe, Pyle, Robinson, Schiller, Smith, Sullivan, Weintraub, & Wepman, 1971; Burke, 1972; Martin & Brogan, 1972; Moss, 1981; Watson, 1973). Such a focus is most commonly manifested by keeping the reader involved with whole, non-fractionated text and also avoiding treatment of aspects or features of written language in isolation.

The Impact of the Teacher's Orientation

Harste and Burke (1977) presented field data to support their hypothesis about theoretical orientation which further suggested that the orientations of many students reflect the orientation of their teacher. Research which has involved the impact of instructional approaches upon reading behavior (Andrews, 1976; Barr, 1975; Biemiller, 1970; DeLawter, 1970; Norton, 1977; Stansell, Harste, & DeSanti, 1978) indicates that reading behaviors are to a considerable degree artifacts of instruction. Since the instructional approaches which teachers use are likely to be reflective of their theoretical orientation (or modified to reflect this orientation), it may be argued that the theoretical orientations held by teachers constitute a major variable in the determination of reading behavior among learners. However, one particular line of research (Bawden, Buike, & Duffy, 1979) has suggested that teachers' theoretical orientations may be only one among many variables influencing their instructional decisions. Specifically, this research indicates that teachers' beliefs about reading may be pushed into background by more immediate "practical" concerns and therefore not function as the major determiner of instructional decisions. Nevertheless, research investigations of the development of theoretical orientations among teacher trainees seem clearly warranted, as do investigations of how such orientations may change.

Questions Guiding the Study

This study was undertaken to address the question of how an undergraduate reading specialization program affects the theoretical orientations of students enrolled in it. Within this general framework, more specific questions include:

1. What is the nature of students' theoretical orientations at the beginning of the program?
2. What are the effects of various campus-based and field-based components within the program upon students' orientations?
3. What is the nature of students' orientations at the end of the program?

METHOD

Organization of the Program

The program in question requires students to complete seven courses, which collectively represent eighteen semester hours of credit. This coursework is followed by a full semester of student teaching. For the purposes of this study, the entire program of coursework-plus-student-teaching was divided into four components, or sections. These four components and a brief description of the courses included in them are given in Table 1.

The division of the program into these four components was not entirely arbitrary. Including the first four courses in Component I was expected to enable the effects of the initial instruction in the teaching of reading to be discernible. The content of these four courses represents that which is typically included in a regular three-semester-hour course on "Reading in the Elementary Grades". Because this content is spread over four courses representing ten semester hours, the coverage is much more extensive and thorough than is possible with one three-semester-hour course. The inclusion of the field-based practicum course alone in Component II reflects an intuitive judgment that field experi-

254

TABLE 1
Components of Program

Component	Description
I.	a. Two courses taken concurrently which are designed to introduce students to basic concepts and terms concerning reading instruction in the primary and intermediate grades. b. A course on the methods, approaches, and materials of reading instruction. c. A course on assessment procedures used in reading instruction.
II.	A field-based practicum course which requires the student to work with children in an instructional capacity for one hour per day over a twelve-week period.
III.	Students may select any two of the following three courses, referred to as 400-level courses. a. A secondary reading course, which is designed for those students interested in teaching in the upper elementary, middle school, or junior high grades, and which treats content area reading from the perspective of the classroom content area teacher. b. A course on the use of children's literature in teaching reading. c. A course on the nature and development of oral and written language process. This course is later referred to as 460.
IV.	Completion of Student Teaching.

Note: The study involved five data collection points which relate to these components:
1. Beginning of Component I
2. Completion of Component I
3. Completion of Component II
4. Completion of Component III
5. Completion of Component IV

ences may tend to produce in students a tendency to reflect the orientations of those practicing teachers with whom they have worked; it was therefore decided to see what, if any, effects could be attributed to this course alone. Component III includes the last two courses to be taken prior to the student teaching experience (Component IV); this partitioning permits comparisons of students' orientations prior to and following student teaching, as well as comparisons of entry-level students, those progressing through the coursework, and those completing the coursework.

Instrument

The instrument, the Theoretical Orientation to Reading Profile (TORP), was developed and validated by DeFord (1978). The TORP consists of 28 propositions about reading instruction, and requires respondents to indicate agreement or disagreement with each proposition on a five-point Likert scale. This instrument is similar to one developed and used by Barr and Duffy (1978) with the important difference being that while their instrument measures preferences for instructional approaches (e.g., basal text, linear skills, natural language, interest-based, integrated curriculum), the TORP measures underlying beliefs about reading and reading instruction. As Harste and Burke's (1978) data indicate, instructional approaches in reading are broadly defined and may therefore accommodate more than one theoretical orientation. However, as these same authors later pointed out (Harste & Burke, 1979), it is the teacher's orientation rather than choice of instructional approach that governs both the specific instructional behaviors in a given classroom and the orientations of the learners.

Among the 28 TORP propositions, there are ten that reflect a phonics orientation, ten that reflect a skills orientation, and eight that reflect a whole language orientation. Scores on the TORP may range from 28–140. Subjects scoring at the low end of this range (28–65) reflect a phonics orientation and those whose scores are in the upper portion of this range (110–140) reflect a whole language orientation. Scores in the middle of the range (65–110) reflect a skills orientation.

Data Collection

In each semester since the Spring of 1979, data have been collected from all those students who had progressed to each of the five data collection points in the program previously described in Table 1. This procedure yielded 761 TORP protocols for analysis. Because data collection proceeded in this fashion for some two and a half years, some students responded to the instrument at several data collection points, while others (e.g., those near completion of the program at the onset of data collection) responded only once.

RESULTS AND DISCUSSION

It was necessary to deal with the fact that at the identified data collection points, students did not all have the same courses completed. For example, some of the students completing Component I had completed additional courses from other components. Therefore, for the purpose of analysis, categories were devised (see Table 2) which reflect the actual courses students had taken at the various data collection points.

The departures from the regular course sequence that were observed among students at the different data collection points were, in a sense, advantageous; the categorization scheme that resulted permitted a closer and clearer examination of the effects of some specific courses than would have been possible if these courses were in every case embedded within their respective components.

The reported comparisons among the TORP scores in various categories must be interpreted with varying degrees of confidence, since sample sizes across categories vary considerably due to several factors. For example, courses in different components vary in their popularity with students; also, as mentioned previously, students deviate from the intended course sequence. In constructing the categories, therefore, an attempt was made to capture the actual course combinations students had put together, and to compare the orientations these combinations produced.

The data were analyzed through the use of descriptive statistics (Table 3) as well as t-tests (Table 4). Inferences that may be drawn from inspection of category means are supported by t-tests comparing means of several categories. As can be seen from Tables 3 and 4, scores of beginning students (category 1) are significantly different from those of students who have finished the first four courses (category 2), from those of students who have finished the first four

TABLE 2
Categories for Comparison*

Category	Description
1	Beginning reading courses (i.e., newly enrolled in the two initial courses of Component I)
2	Finished first four courses (i.e., finished Component I)
3	Finished first four courses and 460, and/or possibly one other 400-level course
4	Finished first four courses and one or more 400-level course *other than* 460
5	Finished first four courses, then 460, then practicum
6	Finished first four courses and practicum
7	Finished first four courses, then practicum and 460 concurrently
8	Finished first four courses, then practicum and two 400-level courses including 460 concurrently
9	Finished all courses including 460
10	Finished all courses not including 460
11	Finished all courses including 460 and finished student teaching
12	Finished all courses did not take 460 and finished student teaching

*These categories are the categories among which comparisons are made and which reflect actual program coursework completed at the points of data collection.

courses and the Language and Reading course (category 3), and from those of students who have finished the first four courses and one of the other 400-level courses (category 4). The t-tests also indicate that the scores of students who have finished the first four courses are significantly different from the scores of students who have also finished the language and reading course, although not from students who have also completed one of the other 400-level courses. However, scores of students in categories 3 and 4 are not significantly different. These results suggest that students entered the program with a phonics orientation and that substantial change ensued as a result of their initial coursework. The results also suggest that the 400-level courses may have produced some changes in orientation beyond those of the first four courses.

Categories 5–8 reflect the coursework configurations of students completing Component II, the field-based practicum course. Whereas the initial coursework in Component I produced higher mean scores, the practicum may possibly exert a slight counterinfluence on the mean scores and score ranges of students who had the Language and Reading course (460) as part of Component I (categories 3, 5), and virtually no influence where students completed only the first four courses prior to practicum (categories 2, 6). Students who took practicum, 460, and another 400-level course concurrently (category 8) tended to generate scores that were comparable to those in category 3, and higher than those in 2 and 6. It appears that the practicum and 460, in this component, may have had counterbalancing effects on students' orientations.

It was also noted that the mean for students completing Component III with

TABLE 3
Descriptive Data from the 12 Categories

Categories*	N	Mean TORP Score	Std Dev	Std Error	Min Score	Max Score
1	217	63.09	9.66	.66	42	108
2	225	85.08	11.60	.77	49	118
3	55	104.40	14.02	1.89	66	133
4	22	93.77	12.52	2.67	63	123
5	5	96.20	6.83	3.06	87	103
6	36	86.28	11.97	2.00	61	110
7	21	99.10	14.42	3.15	78	131
8	12	105.83	16.56	4.78	80	133
9	119	104.33	13.83	1.27	74	140
10	10	86.00	12.27	3.88	67	101
11	37	97.51	14.63	2.41	71	136
12	2	99.00	9.90	7.00	92	106
2,3,4	302	89.23	14.24	.82	49	133
6,7	57	91.00	14.24	1.89	61	131
9,10	129	102.91	14.53	1.28	67	140
11,12	39	97.59	14.34	2.30	71	136

*1 = Entering program; 2 = Finished first 4 courses; 3 = Finished first 4 courses and 460; 4 = Finished first 4 courses and one 400 level course not 460; 5 = Finished first 4 courses, then 460, then practicum; 6 = Finished first 4 courses and practicum; 7 = Finished first 4 courses, then practicum and 460 concurrently; 8 = Finished first 4 courses, then practicum and two 400 level courses concurrently including 460; 9 = Finished all courses including 460; 10 = Finished all courses not including 460; 11 = Finished all courses including 460 and finished student teaching; 12 = Finished all courses, not including 460, and finished student teaching.

the Language and Reading course (category 9, $\bar{X} = 104.33$) was significantly higher than the mean for students completing this component without that course (category 10, $\bar{X} = 86.00$). These differences in means between categories 9 and 10 suggest that Component III makes a difference in students' orientations only if the Language and Reading course has been taken. Indeed, this course appears to exert a strong influence no matter where in the program it is taken. The mean for students in category 9 was higher than the mean for students in all other categories except 3 and 8, which also include 460. Furthermore, the maximum score in category 9 was higher than the maximum score in the other categories, and the means of categories which include the Language and Reading course tended to be higher than the means of other categories.

Finally, the student teaching experience (Component IV) seems to have exerted a slight counterinfluence upon the orientations of students who had completed the Language and Reading course prior to student teaching (categories 9, 11). As mentioned previously, some students responded to the TORP more than once; an inspection of scores from individuals who generated protocols in categories 9 and 11 revealed potentially important differences. While many scores were lower after student teaching than before, some were substantially higher after student teaching and still others were unchanged. Two subjects with identical scores of 109 in category 9 showed a 39-point difference in scores after student teaching. Although there are many potential explanations

TABLE 4
Comparisons Among Categories

Comparisons	T	P>T	Comparisons	T	P>T
1–2	–21.68	.0001	7–8	– 1.22	.2302
1–3	–20.63	.0001	2–9	–12.97	.0001
1–4	–13.78	.0001	2–10	– .25	.8058
2–3	–10.63	.0001	6–9	– 7.07	.0001
2–4	– 3.33	. 001*	6–10	.06	.9488
3–4	3.09	.0028*	7–11	.40	.6923
2–6	– .58	.5658	8–11	– 1.66	.1040
3–5	1.29	.2036	9–11	2.58	.0108*
3–6	6.38	.0001	6,7–9,10	– 5.18	.0001
3–7	1.46	.1477	2,3,4–9,10	– 9.08	.0001
5–6	– 1.80	.0796	2,3,4–6,7	– .86	.39
5–7	.43	.6693	9,10–11,12	2.01	.05
6–7	3.61	.0007	6,7–11,12	– 2.22	.03
3–8	.31	.7571	2,3,4–11,12	– 3.45	.0006

*1 = Entering program; 2 = Finished first 4 courses; 3 = Finished first 4 courses and 460; 4 = Finished first 4 courses and one 400 level course not 460; 5 = Finished first 4 courses, then 460, then practicum; 6 = Finished first 4 courses and practicum; 7 = Finished first 4 courses, then practicum and 460 concurrently; 8 = Finished first 4 courses, then practicum and two 400 level courses concurrently including 460; 11 = Finished all courses including 460 and finished student teaching; 12 = Finished all courses, not including 460, and finished student teaching.

*These differences in category means were not confirmed as significant by the Duncan Multiple Range Test.

for this phenomenon, the possibility that such differences may be due to such factors as ego strength and dogmatism in both student teachers and the cooperating teachers who serve as their mentors in the field should not be ignored.

SUMMARY AND CONCLUSIONS

Based on the assumption that teachers do have a consistent, stable theoretical rationale which underlies and generates teaching behaviors, this study represents an effort to use DeFord's TORP to investigate the effects of a reading teacher education program and its components. Conclusions drawn from the study, which should be regarded as somewhat tentative, are as follows:

1. Introductory courses tend to move students from a phonics orientation to a skills orientation.
2. The field-based practicum alone has no discernible effect on students' orientations. However, it counters the impact of the Language and Reading course if both are taken in Component II.
3. Immediately prior to student teaching, students tend to manifest a skills orientation, unless they have had the Language and Reading course in which case they tend to shift toward a whole language orientation.
4. Overall, the student teaching experience seems to have had little impact on students' orientations; in individual cases, however, orientations appear to have varied somewhat following student teaching.

It must be emphasized that the data in this study arise from a single program and that the impact of various aspects and components of that program is probably in large measure a function of the specific courses and sequences within it. While we would not attempt to generalize these results beyond this program, we would suggest that other such programs do affect the basic conceptualizations of reading among students and ultimately the behaviors of learners in schools. We, therefore, invite others to investigate such programs and to evaluate results in terms of the desirability of the effects they produce.

The data demonstrate the impact of an undergraduate reading specialization program and its specific courses and components. However, an important question concerning this impact and that of teacher education in general involves the influence of the first few years of teaching. It is gratifying to know that an influence on orientation has occurred as a result of course offerings; however, whether that influence is maintained once students are out of the college setting and interacting with a different set of peers is an equally important question which research should address. In addition, the interaction of student teachers and the regular classroom teachers with whom they work needs careful study. Detailed case studies of these interactions could yield valuable insights about those situations in which older, more experienced teachers greatly influence their younger colleagues, and those in which the novice has a real impact on the seasoned professional. Such studies could possibly yield a profile of characteristics which mark present and future leaders in reading education.

REFERENCES

Aaron, I. E., Artley, A. S., Goodman, K. S., Huck, C. S., Jenkins, W. A., Manning, J. C., Monroe, M., Pyle, W. J., Robinson, H. M., Schiller, A., Smith, M. B., Sullivan, L. M., Weintraub, S., & Wepman, J. M. *Scott Foresman Reading Systems*. Glenview, Illinois: Scott, Foresman and Company, 1971.

Andrews, N. C. *Six case studies in beginning reading*. Unpublished doctoral dissertation, Indiana University, 1976.

Barr, R. The effect of instruction on pupil reading strategies. *Reading Research Quarterly*, 1974–75, *10*, 555–582.

Barr, R., & Duffy, G. *Teachers' conceptions of reading: The evolution of a research study*. Research Series No. 17. East Lansing, Michigan: The Institute for Research on Teaching, College of Education, Michigan State University, 1978.

Biemiller, A. The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly*, Fall, 1970, *6*, 75–96.

Burke, C. The language process: Systems or systematic? In R. E. Hodges & E. H. Rudorf (Eds.), *Language and Learning to Read: What Teachers Should Know About Language*. Boston: Houghton Mifflin, 1972, 24–31.

DeLawter, J. *Oral reading errors of second grade children exposed to two different reading approaches*. Unpublished doctoral dissertation, Columbia University, 1970.

DeFord, D. *A validation study of an instrument to determine a teacher's theoretical orientation to reading instruction*. Unpublished doctoral dissertation, Indiana University, 1978.

Harste, J., & Burke, C. A new hypothesis for reading teacher education research: Both the teaching and the learning of reading are theoretically based. In P. D. Pearson (Ed.), *Twenty-sixth Yearbook of the National Reading Conference*. Clemson, S.C.: National Reading Conference, Inc., 1977, 32–40.

Harste, J. C., & Burke, C. L. Understanding the hypothesis, it's the teacher that makes the difference. In K. Vandermuelen (Ed.), *Reading Horizons: Selected Readings*. Kalamazoo, Michigan: Western Michigan University Press.

Martin, B., & Brogan, P. *Sounds of Language Program: Sounds of Jubilee and Sounds of Freedomring*, Teacher's Edition. New York: Holt, Rinehart and Winston, 1972.

McCracken, G., & Walcutt, C. *Basic Reading: Pre-Primer and Primer*, Teacher's Edition. Philadelphia: J. B. Lippincott, 1963.

Moss, R. K. Classroom analysis of teachers' theoretical orientation to reading. *Louisiana Reading Teacher*, Fall, 1981, *4*, 13–22.

Norton, D. *A comparison of the oral reading strategies and comprehension patterns developed by high, average, and low ability first grade students*. College of Education Research Grant No. 15350. College Station, Texas: Texas A&M University, 1977, and Urbana, Illinois: ERIC Clearinghouse ~~on Reading and Communication Skills,~~ 1978.

Robinson, H. M., Monroe, M., & Artley, A. S. *The New Basic Readers: Fun With Our Family and Fun Wherever We Are*, Teachers Edition. Chicago: Scott, Foresman and Company, 1962.

Stansell, J., Harste, J., & DeSanti, R. The effects of differing materials on the reading process. In P. D. Pearson (Ed.), *Twenty-seventh Yearbook of the National Reading Conference*. Clemson, S.C.: National Reading Conference, Inc., 1978, 27–35.

MICHAEL L. KAMIL
University of Illinois at Chicago

TECHNOLOGY AND READING: A REVIEW OF RESEARCH AND INSTRUCTION

There is an historic and continual tension between technology and the arts. Some argue that our culture actually places art and philosophy at the center of human existence with science and technology at the periphery (Burke, 1978). Simon (1981) on the other side of the question argues:

> Human behavior is a supremely articificial phenomenon . . . As long as we avoid physical or biological damage, however, we are capable of an immense range of alternative behaviors, only a fraction of which appear to be wired in.
>
> One important way, then, in which human beings evade the limits of intellectual performance imposed on them by their "hardware" is by using external aids to memory. These supplement both STM (e.g., paper and pencil) and LTM (e.g., accounting records, encyclopedias), thereby enhancing tremendously the thinking of which human beings are capable.
>
> A second way in which human beings overcome the limits of their information-processing equipment is by learning to use that equipment more efficiently—by discovering and adopting more efficient strategies for accomplishing thinking tasks.

Both attempts to evade human limits are apparent in the effects of technology on reading research and instruction.

Singer, Dreher, & Kamil (In press) give a time-line for the development of technology as it relates to the classroom. Paper, pencils, and alphabetic writing were developed around 3000 to 4000 years ago. Chalk and blackboards appeared between then and the development of the Gutenberg printing press in 1440. A copy of the Mainz Psalter survives today that was printed with movable type in 1457. Photography was invented in 1839; the phonograph was invented in 1878. Marconi invented the wireless radio in 1895. The DeForest vacuum tube dates from 1907. Television was invented in the 1930s and became widely available in the 1950s. The spread of television has been phenomenal. By 1977, 72% of all teachers had television available, even though only 59% used it instructionally.

Computers are very recent developments. The Mark I, the first automatic digital computer, was invented in 1939; ENIAC, the first electronic computer, was developed in 1945. ENIAC weighed 30 tons, occupied 3,000 cubic feet of space and used about 18,000 vacuum tubes, which blew out about once every 7½ minutes. Early in the 1970s, the first microcomputers were developed. They weighed about 30 pounds, fit on a desk top, were faster, easier to program, and cost under $500. If automobiles had benefitted from technology at the same rate, a Rolls Royce would cost about $2.75, get about 3,000,000 miles to a gallon of gasoline, and have the power to propel an ocean liner (Evans, 1979, p. 77).

Overview

This review will concentrate on the application of computers to reading research and instruction. Reviews of other aspects of technology can be found for reading machines and tachistoscopic devices in Harris & Sipay (1980), for a variety of devices in Kamil, Moe & Johnson (In press), Moe & Johnson (1980) and for television in Mason & Mize (1978). This is not an exhaustive list.

It has been a long time since the last review of technology appeared in an *NRC Yearbook* (Spache, 1967). In his review, Spache identified three areas of technology that held promise for reading research and instruction. The first of these was the use of O. K. Moore's talking typewriter. At present, there seems to be little interest in the talking typewriter. It never seemed to fulfill the promise it had in 1967.

Another area was programmed instruction. The origins of programmed instruction are to be found in Pressey's development of a machine to score multiple-choice exams in 1924. Major development of the techniques is attributed to Skinner and his colleagues at Harvard in the 1950s (Skinner, 1958). Many of the present uses for programmed instruction are intimately bound up with computers and computer-assisted instruction (CAI).

The final effects of technology which Spache identified were the uses of computers to do tasks like letter counts, readability calculations, determinations of occurrences of sentence patterns, and linguistic development simulations. These have all been done with more or less success. Even syllable counts have been compiled and published (Sakiey, 1979). The word lists of Carroll, Davies and Richman (1971) or Kucera and Francis (1974) could not have been compiled and manipulated so easily if it were not for computers. (However, it is important to note that the Thorndike & Lorge (1944) lists were generated without the benefit of computers.)

Technology and Reading Research

There are three main contributions of technology to research in reading. The first is simply in data collection. Data about reading processes can be obtained from a range of instrumentation from simple to complex. However, some of the most sophisticated data could not be obtained unless appropriate technology existed for its collection. A primary application of technology is the study of eye movements begun by Javal, Dodge and other early researchers (c.f. Huey, 1908) and continuing to the present in the work of such researchers as Fisher, McConkie, Rayner, Just, and Carpenter. A thorough review of most of this work can be found in Rayner (1978). Other areas where technology is crucial are the collection of reaction time data, physiological data, EEG data, hemispheric response data, or the use of video tapes to do descriptive or observational studies. Not all of these examples require the use of computers; some are simply more precise when computerized.

Not only is the instrumentation available to collect data about the reading process, but automated data collection is possible through computer control. Large numbers of subjects can be used in the same experimental setting. Multiple observations can be made on each of those subjects, with automatic recording and processing of the data. Further, the computer can tailor stimulus patterns in any of innumerable manners, from random to fully dependent on the prior responses of the participants.

With concerns about ecological validity (Bronfenbrenner, 1978) researchers can turn to computers to allow them to collect sufficient data to describe the experimental environment in a more nearly complete manner. It is only with a computerized data collection system that this becomes remotely possible. Subsequently, the computer could be used to recreate the same conditions under which the first observations had taken place—or to measure the environment in sufficient detail to indicate how the two situations differed.

A second effect that computers have had on research is in the analysis of the data once it is collected. It is hard (but not impossible) to imagine performing multiple regression or multi-factor analyses of variance by hand. While it is possible, the time involved is clearly prohibitive. Thus, only computerized analysis makes it reasonable to use effectively the large amounts of data that can be collected under automated control. (Note there is no inherent guarantee of quality in the collection of large amounts of data—computers only provide the opportunity to collect and analyze data. Humans must still evaluate the statistical results. An old computer slogan applies here: GIGO—Garbage In; Garbage Out.)

At one time the analysis of data by computer required a great deal of effort and technical hardware knowledge of computers. The development of high-speed computers and sophisticated software has changed that. Almost every reading researcher has used SPSS (Statistical Package for the Social Sciences) to analyze some sort of data on a computer. New developments are making that task even easier. A new program, SCSS (Statistical Conversations for the Social Sciences) requires only that the user be able to answer questions regarding the design, variables, number of observations, etc. There is no need to format responses, write key-words or the like. Future developments may reduce even this minimal requirement.

The final category of effects of technology on reading research is in the use of technology to model the reading process. It is clear that models such as that of Gough (1972), LaBerge and Samuels (1974) or Rumelhart (1979) would not have had their present forms were it not for the use of computer programming and/or simulation techniques. The work of Kintsch and van Dijk (1980) shows how computer simulations have effects on the production of models of text processing. In a different vein, Carpenter and Just (1980) have used computer-collected eye movement data to develop a model of the reading process.

The entire line of work that has come to be called "schema theory" in reading has been largely shaped by computer modelling and simulation. The work of Schank and Abelson (1977) illustrates much of what can and has been done with computer simulations of comprehension.

Future developments in applications of computers to reading research will be limited only by the imagination of the researchers. Even now, the availability of microcomputers to almost all researchers has an impact on the quantity and quality of reading research. More sophisticated hardware will produce an even broader range of possibilities. There will also be further noticeable effects in modelling and simulations as more reading researchers gain access to computers. Microcomputers now have the capabilities to run some of the more sophisticated languages used for simulations, like LISP. The use of networks to link microcomputers to each other or to central information banks may facilitate the communication among researchers and thus produce more cooperative efforts. Finally, the ultimate availability of information retrieval services by computer will provide easy and instant access to datasets, journals and other materials for both submission and distribution. As only a single example, computer tapes are available from the National Assessment of Educational Progress so researchers can perform their own analyses on the coded test data.

Technology and Reading Instruction

This section will deal primarily with the effects of computers on reading instruction. However, other technological effects deserve mention. A large body of literature exists on the effects of machines on reading rate. Included in this are tachistoscopes, reading pacers, and eye movement analysis devices. A good review of these machines can be found in Harris and Sipay (1980).

The work of Sticht (1974) has shown some improvement in reading comprehension as a function of auding training. While speech compression devices exist and seem to have some potential for reading applications, there has been limited use of these devices in reading. The work of Carver (1977–78) has also shown the connection between reading and listening.

Materials preparation has been greatly affected by technology. The use of alternative media, word processing, improved printing techniques, and duplicators has extended the range of materials available to individual instructors. The phenomenon of mass production is a major result of technology. At the same time, educational management systems have depended in large measure on the development of technology—computers, high speed scorers, etc. Many management systems have also been conceptualized to allow them to make use of technology. Skills, behavioral objectives, and the like are easily adaptable to machines. Many of the skills management systems depend on computers for either record-keeping or prescription. Several companies presently market microcomputerized management and prescription systems. Other systems are available or under development to accompany specific basal reading series.

Television and radio have had considerable impact on teaching reading. There are other reviews of these areas, so they will not be dealt with here. (One good source is Mason and Mize, 1978.)

Teaching and learning machines have had effects on reading for some time. Similar devices can be found in the current collection of computer-like machines on the market (e.g., Speak and Spell, Speak and Read, etc.). A review of some of these alternatives to computers can be found in Anderson (1981).

Another device that is now available is the Kurzweil reading machine (Cushman, 1980). This machine can translate text to sound with a high degree of accuracy. It is text independent. That is, it can be used without special preparation of the materials to be read. However, the cost has limited the use of these machines to those who require them because of a handicap. While it is highly accurate, it can "mis-read" when presented with odd-sized type or unusual fonts.

Finally, the area of computer instruction and computer literacy needs to be considered. Under this heading come both instruction and word processing. Word processing is presently a major use of computers for business. It represents a potentially large occupational area. It will probably be a major influence in the future of higher education as it becomes the dominant mode for writing. At the University of Illinois at Chicago, most students have access to word processing systems on the university computer as a routine. Theses and term papers are routinely prepared on the system. Faculty also use it for preparation of publications, course materials, and other professional materials. This is only one example; the University of Illinois is not unique in this regard.

Computers and Literacy

The remainder of this review will be concerned with CAI and computer literacy and their effects on reading instruction. Perhaps the most visible effect of technology on reading instruction has been its effect on the definition of literacy. It was once the case that being able to sign one's name was a sign of literacy. A combination of technology and military and other large-scale manpower needs have re-defined literacy. It was only recently that literacy came to mean the ability to read and comprehend new material (Resnick & Resnick, 1977). Today, literacy often includes the abilitly to interact with computers—computer literacy. However, even this will require a changing definition. While ambitious programs to develop computer literacy involved knowledge of programming and the like (c.f. Johnson, Anderson, Hansen & Klassen, 1980) further technological developments may make some forms of computer literacy unnecessary. For example, computer literacy for early computers required knowing how to wire program boards by hand. Later machines could be used only if one could do the key punching of cards. Today, microcomputers rarely have the option of using cards for input. Present-day, state-of-the-art programs for writing CAI/tutorial materials require almost no programming knowledge. Users interact with the computer in (almost) conversational fashion.

As an analogy, we do not expect people to repair a washer or dryer although they can easily operate them. Beginning drivers are not expected to know auto mechanics. Television viewers would be hard-pressed to explain how their sets worked. Computer literacy may become simply knowing how to use computers, not necessarily how to program them in detail.

What has become vitually important is information. Computers handle information better than other devices, but it is important not to lose sight of the fact that books are highly efficient random access devices. That is, any information can be accessed without dealing with other information. Tape requires sequential operations in which every preceding piece of information must be covered (or ignored). Only videodiscs come close to approximating the efficiency of books as random access devices. In short, computers are the first technological advances that even come close to being true substitutes for books as storage devices for information.

Reading and CAI

The first major efforts to use computers in teaching reading were those of the Brentwood Project, later called the Stanford Project under the direction of Atkinson, Suppes, and their colleagues. A report on the benefits of CAI for reading created some negative reaction. (See Atkinson & Hansen, 1966 and comments by Spache, 1967b and reaction by Atkinson, 1967.)

A typical lesson was divided into six parts (Atkinson & Hansen, 1966): letter discrimination, initial vocabulary acquisition, word decoding tasks, syntactic and intonation practice with phrases and sentences, syntactic and semantic practice with phrase and sentence material, and information processing tasks. It was not the general categories that evoked reaction, but the fact that the computer did little more than allow for drill and practice on relatively isolated skills (Spache, 1967b).

It should be noted that the original project used very large computers, with teletype displays and audio presentations through headsets. Audio communication was uni-directional. This project was the basis for some current efforts to teach reading with computers. Computer Curriculum Corporation provides commercial services for schools who want entire reading (and other) curricula delivered by computer. While the present Computer Curriculum Corporation materials and hardware are very different from those used originally, they can be traced back to the early work.

There have been several large-computer attempts to develop CAI for reading and other curriculum areas. Most notable among these has been the PLATO (Programmed Logic for Automatic Teaching Operation) project originally developed at the University of Illinois and now largely based at the University of Delaware and Control Data Corporation. PLATO offered several advantages over other computer systems: A special terminal, with superior graphics, touch panels for simplified responding, special easy-to-use languages for programming instructional materials and random access audio output. PLATO has been used to teach beginning reading, college level English, and GED reading comprehension, among a very large list of other subjects.

Another major system is the TICCIT system (Time-shared Interactive Computer-Controlled Information Television) developed by Hazeltine and Mitre Corporation using Data General computers. Much work has been done with TICCIT at Brigham Young University. The unique advantages of TICCIT are a special keyboard to make input easier and a display that incorporates television video tapes, normally or frame by frame. Other large-computer applications in reading are described in Mason & Blanchard (1979).

Advantages of CAI

There are, theoretically, major advantages to CAI. Instruction can be individualized with regard to rate and content. A totally unique curriculum (in terms of actual content) could be developed for each student, given the extremely large number of difference choices a computer program might make based on student performance. Attention and motivation can be built into CAI fairly easily. The "addiction" to video games is often utilized in instructional environments for such purposes. Computer programs are objective; they are unaffected by factors not directly relevant to learning. Computers are programmed to make instructional decisions on the basis of academic performance. Often teachers do not attend to relevant variables when they have to make decisions in the classroom (Duffy, 1981). Finally, some tasks can't be done by humans—clear examples are most readily found in mathematics. An iterative solution to a complex mathematics problem can be found in a few seconds or minutes with a computer; it might involve a lifetime of work for a mathematician.

Problems With CAI

Despite the advantages, there has been less than stellar progress toward realizing them. The major block has traditionally been the expense. Large computer systems for CAI were beyond the fiscal reach of most school systems. Even time-sharing on large systems did not reduce the cost sufficiently. It has only been in the last few years that microcomputers have been inexpensive enough for most schools to be able to afford them. Further reductions in cost will be needed before computer technology will be truly affordable in schools. It may be, however, that we can't afford *not* to afford computers. Schools must prepare students for the impact of computers on their lives.

Additionally, large computer systems have a major fault: They often go "down". When a large computer is out of commission, all of the terminals connected to it on a time-sharing basis are also out of commission. Again microcomputers have eliminated some of this problem. Each stands alone, if one goes "down" it does not affect the others.

Mason (1980) lists as additional problems, the cost of software, the difficulty of reading at a CRT (cathode ray tube), the fact that a computer cannot listen to someone read in order to help improve his/her fluency, and that the computer is limited to the responses that are programmed.

The software problem is indeed severe. The following calculations illustrate some of the problem. Assume that we are to provide CAI for all 12 grades. For each curriculum area, there is one textbook per grade with about 10 chapters, making 120 chapters per grade. About five 8K programs are the equivalent of one chapter. This yields a figure of about 600 programs per curriculum area. About 6000 programs would be required to computerize ten curriculum areas. If we consider that we would need only four different program versions, one for each of four major microcomputer manufacturers (e.g., Atari, Commodore, Radio Shack, and Apple), we would need 24,000 programs. A cost of $8000 is used for an individual program, giving an estimate of $192 million for software alone (Isaacson, 1981). This may be a lower-bound estimate; the more flexible and capable the software, the more expensive it will be.

Cost is not the only software problem. Who is to do the authoring? Who should decide where computer materials fit in the curriculum? How should software be evaluated? Who should do the evaluation? Should computer software be correctable at local sites? All of these questions and more need to be addressed. A part of the reason that CAI did not become more popular in the 60s was the lack of what might be called "humanistic" software. It will take a concerted and cooperative effort on the part of reading professionals, computer programmers, and publishers to make high quality materials available for use with microcomputers. Moreover, the materials will have to make use of the special capabilities of computers to be widely accepted by reading professionals.

Research on Reading and CAI

Most of the studies that have been done have involved larger computer systems. They are not exactly relevant to the present state of affairs where there are far larger numbers of microcomputers (which have been commercially available since about 1974–75) in use than there were when most of the studies were conducted. In light of such a strong caveat, what do the results have to suggest about CAI?

Vinsonhaler & Bass (1972) surveyed the literature on drill and practice in CAI. They found an advantage of .1 to .4 school years for CAI drill and practice groups in language arts. Edwards, Norton, Taylor, Weiss & Dusseldorp (1975) did a similar survey, but did not restrict their search to drill and practice. Of the six studies cited, none in reading or English showed negative results. Only one showed equal effects of traditional instruction and CAI; five showed advantages for CAI.

Jamison, Suppes & Wells (1974) compared the effects of many alternative instructional media. For programmed instruction and CAI they reported that there were many findings of no differences. They did suggest, however, there could be some savings in time. Most important, they found that small amounts of CAI, when used as supplements, led to improvements in achievement, particularly for slower students. They also indicated that the overwhelming majority of efforts in developing educational technology were directed toward improving quality with little regard to cost.

Kulik, Kulik & Cohen (1980) have done a meta-analysis that shows the benefits of CAI at the college level. There are consistent, though small, benefits in achievement and affect as the result of CAI. The authors also report a substantial savings in time when CAI is used. Walberg and Vinsonhaler are in the process of doing a similar analysis for CAI at the elementary and secondary school levels (Walberg, personal communication, 1981).

The effects of CAI have been summarized as follows:

1. CAI can be used successfully to assist learners in attaining specified instructional objectives.
2. There is substantial savings in time (20%–40%) required for learning over traditional instruction.
3. Retention is at least as good if not superior to retention following conventional instruction.
4. Students react positively to good CAI programs; they reject poor ones (Gleason, 1981).

Contemporary Research on CAI

Gleason (1981) also points out that there are few researchers who are asking questions of the form: Is CAI better than method X? There are researchers and practitioners working on the following list of problems and questions:

1. What are the most effective CAI strategies? What is best in terms of feedback? Are different paradigms better for different areas?
2. How do individual learning styles interact with CAI? How much cognitive complexity can students handle?
3. What are the effects of individual learner characteristics in CAI situations? (E.g., memory span, perception, IQ, motor skills, sensory preferences, and literacy levels.)
4. What hardware configurations are most effective, efficient, or necessary? How important are audio outputs, touch panels, video capabilities, or light pens?
5. How does affect relate to CAI? What are the roles of motivation, persistence, delayed gratification or locus of control?
6. What are the most effective strategies for program development? How should programmers and practitioners relate to each other?
7. What are the most effective strategies for integrating CAI with other instructional activities?

The answers to these questions will shape the future role of computers in reading instruction. Presently, we have few of the answers; we need many of them.

Computers and Writing

While this is a review of reading and technology, one of the most intriguing areas of application has been in the use of computers for writing. Since some current thought favors the holistic teaching of reading and writing, these applications are important.

Descriptions of the story maker technique can be found in Rubin (1980). Adaptations of that technique for computers are described in Rubin (1982).

Another promising use of computers has been to allow students to compose on the computer with a word processor. Much of the burden of physical corrections are removed, and students can make revisions as often as possible. A package of programs which will include a system for electronic messages is under development by Bruce, Rubin and others (Bruce, personal communication).

These seem to be only the beginning of extensive applications in the productive language arts. We can look for composition and writing to be done on computers routinely, as the re-integration of reading and writing takes place in the curriculum.

Speculations About the Future

No review of this sort is complete without some serious speculation about what the future will hold. The premise of futurists such as Toffler (1972) is that technology is developing at exponential rates. If that is true, the task of prediction is close to futile. However, a few guesses are still in order.

There will be a new definition of literacy. Being an educated person will involve knowing how to interact with computers. It will probably become impossible to complete school without being able to deal with computers at some level. Most students will learn how to type and program as basic skills.

An extensive array of hardware will be available. Computers will eventually have superior auditory output and will have speech recognition capabilities. The machines in the near future will have highly improved graphics, extensive memory capacity, and easy access to video tape/disk input and output. Costs will probably decrease substantially; capabilities will be greater.

Computers will be used to teach all sorts of skills from creative writing to art and music. There will be less drill and practice (in the present sense) in computer software. The choices for learning will not be location-specific. For all practical purposes, the choices of subject matter will become unlimited. A content specialist in each area will not have to be present at each site for students to learn that content.

Students will be able to make use of computers for word processing, learning to write, preparing of materials for class projects, and for interacting with information banks of all sorts.

A new form of "literacy" may appear. Even now, computerized video disk access is being used instead of some training manuals, or even medical diagnosis. For some tasks, it is easier to show a picture of the performance than it is to describe it. For example, it is probably easier to show someone how to tie a bowline knot than to tell them how to tie it. The reading requirements for certain tasks would certainly be minimized.

There will be extensive networking of microcomputers among schools, in manners similar to that used by the Minnesota Educational Computer Consortium. Software development will be a cooperative venture among teachers, programmers, researchers and publishers. Libraries will have computer access to large information services and eventually, most printed material will be available from such computer banks.

Schools could become decentralized. Learning will not have to take place at a single location. However, this will probably happen in the continuing or adult education areas first. It will be possible for teachers to go back to more basic teaching: social and moral development, civic responsibility, creative thinking and the like.

Conclusion

Whatever the future holds, there will be more computers with more sophistication doing more tasks. Literacy will have to include the ability to interact with computers. Reading will still be required, but with different emphases. The promise of technology is great at present, but there are dangers. As James Burke (1978) writes:

> The key to success will be the use we make of what is undoubtedly the vital commodity of the future: Information . . . unless changes are made in the way information is disseminated, we will soon become a society consisting of two classes: The informed elite and the rest. The danger inherent in such a development is obvious.

REFERENCES

Anderson, B. Inexpensive alternatives to computers for classroom use. *Reading Psychology*, 1981, *2*, 103–106.

Atkinson, R., & Hansen, D. Computer-assisted instruction in initial reading: The Stanford project. *Reading Research Quarterly*, 1966, *2*, 5–25.

Bronfenbrenner, U. The experimental ecology for education. *Educational Research*, 1976, *5*, 5–15.

Burke, J. *Connections*. Boston: Little, Brown, 1978.

Carrol, J., Davies, P., & Richman, B. *American heritage word frequency book*. Boston: Houghton, Mifflin, 1978.

Cushman, R. The Kurzweil reading machine. *Wilson Library Bulletin*, 1980, *22*, 311–315.

Duffy, G. Teacher effectiveness research: Implications for the reading profession. In M. Kamil (Ed.), *Directions in reading: Research and instruction*. Washington, D.C.: The National Reading Conference, 1981.

Edwards, J., Norton, S., Taylor, S., Weiss, M., & Dusseldorp, R. How effective is CAI? A review of the research. *Educational Leadership*, 1975, *33*, 147–153.

Evans, C. *The Micro Millenium*. New York: Washington Square Press, 1979.

Frederick, F. *Guide to microcomputers*. Washington, D.C.: Association for Educational Communication and Technology, 1980.

Gleason, G. Microcomputers in education: The state of the art. *Educational Technology*, 1981, *21*, 7–18.

Gough, P. One second of reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and eye*. Cambridge, Mass.: MIT Press, 1972.

Harris, A. & Sipay, E. *How to increase reading ability*, seventh edition. New York: Longman, 1980.

Huey, E. *The psychology and pedagogy of reading*. New York: Macmillan, 1908.

Isaacson, D. What's holding back computer use in education? *Classroom Computer News*, 1981, *1*, 1 and 28–29.

Jamison, D., Suppes, P., & Wells, S. The effectiveness of alternative instructional media: A survey. *Review of Educational Research*, 1974, *44*, 1–67.

Johnson, D., Anderson, R., Hansen, T., & Klassen, D. Computer literacy—what is it? *Mathematics Teacher*, 1980, *73*, 91–96.

Johnson, D., & Moe, A. Current approaches, part two. In P. Lamb & D. Arnold (Eds.), *Teaching reading*, second edition. Belmont, California: Wadsworth, 1980.

Just, M. & Carpenter, P. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 1980, *87*, 329–354.

Kamil, M., Moe, A., & Johnson, D. The hardware and software of developmental reading. In F. Conner (Ed.), *Handbook of developmental reading*. New York: McGraw-Hill, In press.

Kintsch, W., & van Dijk, T. Toward a model of text comprehension and production. *Psychological Review*, 1978, *85*, 363–394.

Kucera & Francis. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.

Kulik, J., Kulik, C., & Cohen, P. Effectiveness of computer-based college teaching: A meta-analysis of findings. *Review of Educational Research*, 1980, *50*, 525–544.

LaBerge, D., & Samuels, J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293–323.

Mason, G. Computerized reading instruction: A review. *Educational Technology*, 1980, *20*, 18–22.

Mason, G., & Mize, J. Teaching reading with television: A review. *Educational Technology*, 1978, *18*, 5–12.

Mason, G., & Blanchard, J. *Computer applications in reading*. Newark, Delaware: International Reading Association, 1979.

Moe, A., & Johnson, D. Current approaches, part one. In P. Lamb & D. Arnold (Eds.), *Teaching reading*, second edition. Belmont, California: Wadsworth, 1980.

Rayner, K. Eye movements in reading and information processing. *Psychological Bulletin*, 1978, *85*, 618–660.

Resnick, D., & Resnick, L. The nature of literacy: An historical exploration. *Harvard Educational Review*, 1977, *47*, 370–385.

Rubin, A. Making stories, making sense. *Language Arts*, 1980, *57*, 285–298.

Rubin, A. The computer confronts language arts: Cans and shoulds for education. In A. C. Wilkinson (Ed.), *Classroom computers and cognitive science*. New York: Academic Press, 1982.

Rumelhart, D. Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, N.J.: Erlbaum, 1977.

Sakiey, E. The commonest syllables. In M. Kamil & A. Moe (Eds.), *Reading research: Studies and application*. Clemson, South Carolina: The National Reading Conference, 1979.

Simon, H. Studying human intelligence by creating artificial intelligence. *American Scientist*, 1981, *69*, 300–309.

Singer, H., Dreher, M., & Kamil, M. Technological innovations in secondary schools: From language inception to computer literacy. In A. Berger and H. A. Robinson, *Secondary school reading: What research reveals for classroom practice*. In press, National Council for Research on English.

Skinner, B. F. Teaching machines. *Science*, 1958, *128*, 969–977.

Spache, G. A reaction to Computer-assisted Instruction in Initial Reading: The Stanford project. *Reading Research Quarterly*, 1967, *3*, 101–109. (a)

Spache, G. Reading technology. In G. Schick & M. May (Eds.), *Junior college and adult reading programs—expanding fields*. Milwaukee, Wis.: The National Reading Conference, 1967. (b)

Sticht, T., *et al. Auding and reading: A developmental model*. Alexandria, Va.: Human Resources Research Organization, 1974.

Thorndike, E., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College Press, 1944.

Toffler, A. *Future Shock*. New York: Random House, 1970.

Vinsonhaler, J., & Bass, R. A summary of ten major studies on CAI drill and practice. *Educational Technology*, 1972, *12*, 29–32.

CHRISTINE J. GORDON
CARL BRAUN
University of Calgary

STORY SCHEMATA: METATEXTUAL AID TO READING AND WRITING

Schema theorists postulate the existence of "mind macrostructures" or internalized story schemata which potentially act as a framework to organize information during the reading act and as retrieval cues for the recall of information bound to the story structure (Kintsch, 1975; van Dijk, 1977). Several similar formalisms, dubbed story grammars, have been developed to represent the macrostructures which underlie the organization of narrative text (Mandler and Johnson, 1977; Stein and Glenn, 1977; Rumelhart, 1978; Thorndyke, 1977).

Researchers (Brown, 1975, 1976; Stein and Glenn, 1978; Mandler and Johnson, 1977) have shown that categories vary in terms of their salience in recall depending on conditions and/or procedures used to test comprehension (free recall immediately after oral presentations, recognition tasks and question probes). These findings prompted Chodos and Mosenthal (1978) to focus on reading rather than listening comprehension and to investigate whether children's story schemata changes under different test conditions (immediate written recall and immediate forced recall requiring students to fill in deleted information in statements). Their findings on the free and forced recall conditions provide support for the notion that children employ schemata in understanding written stories much as they do in understanding spoken stories and over several task demands.

Studies have shown that incidental learning of aspects of these macrostructures transfers to recall of similarly organized selections (Bower, 1974, 1976; Thorndyke, 1977). More recently the question of the effect of direct instruction of story structure has been a subject of investigation. A study by Gordon (1980), for example, addressed itself to the question of whether story structure can be taught as a metatextual aid to reading comprehension. Fifth grade students were taught to apply a simplified story schema to basal reader stories. A macrocloze technique was used. As the stories were read, students filled in certain slots on the story outline with information from the story. Results showed a significant increase in recall of high level information in the story summaries they wrote (specifically in the categories minor setting, initiating event, reaction, and final resolution). But, the facilitory effect of text structure as an aid to comprehension was not evident on tasks that required open-ended responses to wh- probes. Further, no transfer stories were read and summarized by the treatment groups. Also, what was not examined in the Gordon study was the acquisition of story schema as a transferable scaffolding for generating creative writing productions.

The major purpose of this study was to investigate the effects of story schema training on reading comprehension and writing. More specifically, it was hypothesized that story schema would serve as a transferable framework for storing and retrieving information from new and unfamiliar selections read, a framework for answering literal and inferential question probes, and a cue system for generating information in the writing process. The first two predictions were that direct instruction in story structure would increase the number of text structure categories generally found lacking in children's recall after a narrative used in instruction, and that such instructional effects would transfer to the reading of an unfamiliar but similarly organized narrative. The third prediction was that story schema (structure schemata) awareness would contribute significantly to literal and inferential comprehension if prior knowledge on the topic (content schemata) is developed to the same extent in both treatment groups. Finally, children will utilize their knowledge of story schema to generate information in independently written productions.

METHOD

Subjects

The 57 subjects for the study were drawn from three fifth grade classes from one school population. These classes were randomly assigned to experimental and control groups.

Materials

The experimental test materials consisted of three fables adapted to include an equal number of story structure categories. For two other fables, short scripts were written to provide subjects with background content schemata requisite for inferencing from the fables. Wh- probes were constructed for the two fables. These probes were tied to specific text structure categories. An additional constraint was that probes included textually-explicit, textually-implicit and scriptally-implicit categories (Pearson and Johnson, 1978).

Both the inclusion of specific text structure categories and the taxonomic probe categories were validated by three independent raters.

Similarly organized legends and fables were chosen as instructional selections. Where necessary, minor modifications were made to have the stories correspond to a similar organization. The macrostructure of each narrative was then determined based on an adaption of Stein and Glenn (1977) and Thorndyke (1977). Three separate raters validated the categorization of statements in each narrative.

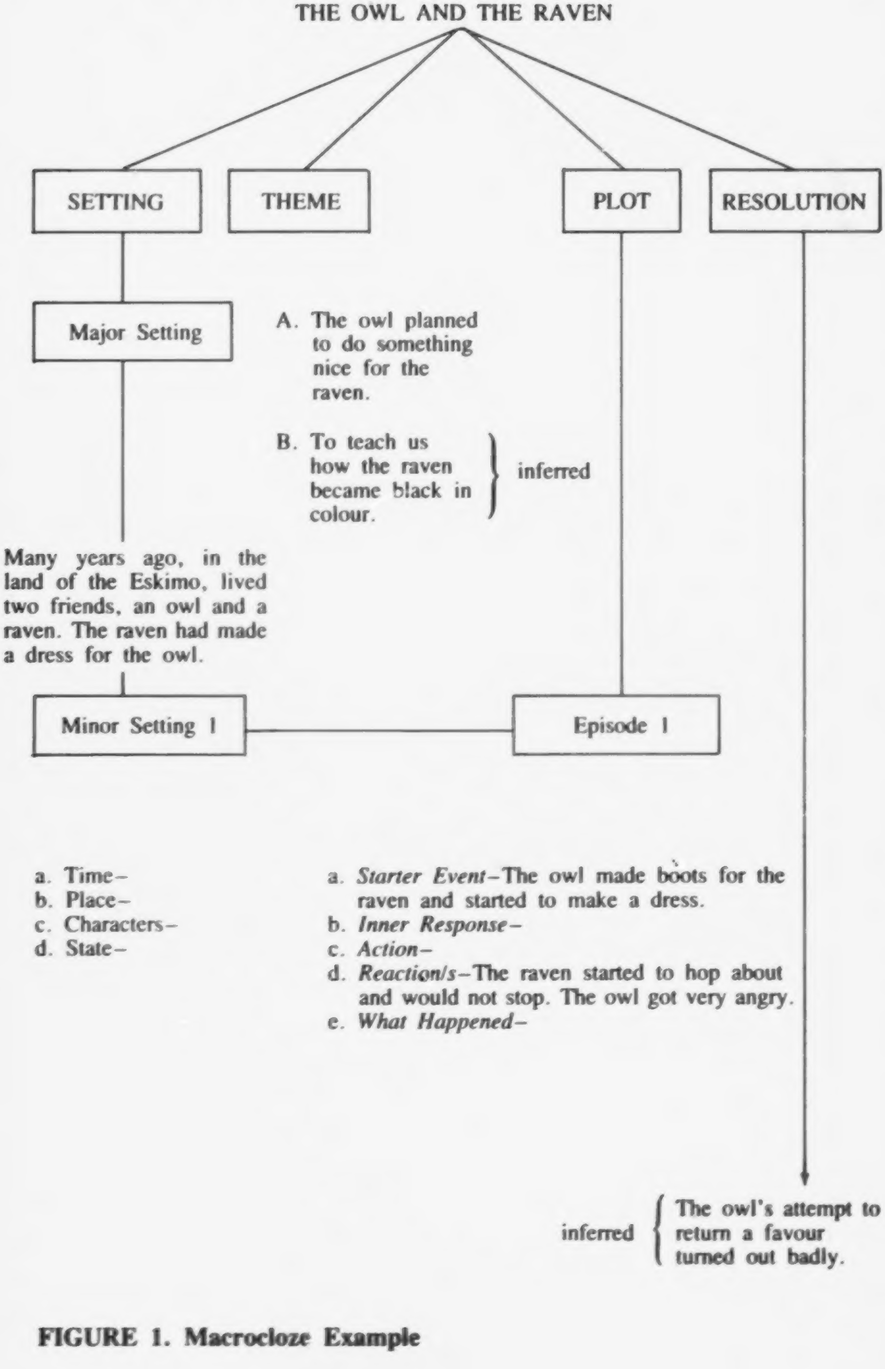
Instructional Method

Initial instruction consisted of exposure to a simplified, global organization structure of a narrative. Macrocloze procedures were used, i.e., templates were presented with some text structure categories containing specific story content while other categories were left open for children to complete. "A priori" questions were used to guide the search for information contained in specific text structure categories. Upon consensus, such information was written by the teacher/investigator on the macrocloze template (see Figure 1). The use of an inductive approach was used to deal with specific categories in three similarly organized fables before proceeding to all text structure components in one narrative.

Succeeding lessons guided students in posing their own questions prior to reading unfamiliar narratives.

In an attempt to enhance transfer effects, the remaining sessions involved the writing of narratives, first with teacher-guidance and then without. These narratives were discussed in terms of text structure components.

The control group received teacher-directed discussion, probes and activities



THE OWL AND THE RAVEN
SETTING
THEME
PLOT
RESOLUTION
Major Setting
Minor Setting 1
Episode 1
inferred
inferred

A. The owl planned to do something nice for the raven.

B. To teach us how the raven became black in colour. } inferred

Many years ago, in the land of the Eskimo, lived two friends, an owl and a raven. The raven had made a dress for the owl.

a. Time–
b. Place–
c. Characters–
d. State–

a. *Starter Event*–The owl made boots for the raven and started to make a dress.
b. *Inner Response*–
c. *Action*–
d. *Reaction/s*–The raven started to hop about and would not stop. The owl got very angry.
e. *What Happened*–

inferred { The owl's attempt to return a favour turned out badly.

FIGURE 1. Macrocloze Example

related to drama and literature appreciation. In order to offset any biasing effects of the experimental writing tasks, the control was engaged in the writing of drama.

The instruction for both experimental and control groups consisted of 15 sessions of 30 minutes duration spaced regularly over a five-week period.

Dependent Variables

The dependent variables used for pre- and posttests were: a) total number of text structure categories produced in written recall of an unfamiliar narrative; b) number of correct responses to wh- probes following the reading of an unfamiliar narrative; c) number of text structure categories included in a creative writing production of a narrative for which a one-sentence "starter" had been provided. An additional posttest measure consisted of the total number of text structure categories produced in written recall of a narrative used in instruction.

Scoring

Propositions in subjects' written recalls were compared to validated macrostructures of each narrative and categorized into specific text structure categories. Propositions in the creative writing productions were classified into specific text structure categories. A set of criteria was also used to score the open-ended responses to the wh- comprehension questions. Three independent raters were involved in the scoring procedures.

RESULTS

While experimental and control groups were randomly assigned, assignment of subjects within groups was not random. Significant differences in pre-measures between groups was found on both intelligence (Otis-Lennon) and reading measures (Canadian Tests of Basic Skills). Multivariate analyses of covariance were, therefore, used to compare the experimental and control groups. No significant differences between experimental and control groups were found on the pretest measures administered to obtain baseline data (written recall of an unfamiliar selection, wh- probes, and a creative writing task). No significant differences between groups were found on the total number of text structure categories produced in the written recall of an unfamiliar selection, the independent writing task or on the total number of wh- probes answered correctly on an unfamiliar fable ($F\ (3,51)=0.18$, $p<.91$).

The written recall, wh- comprehension probes, and creative writing productions obtained after intervention were analyzed using a multivariate analysis of covariance with group as a factor and intelligence, standardized reading, and corresponding pretests as covariates. A multivariate effect for group was obtained ($F\ (3,48)=5.069$, $p<.004$). Separate univariate analyses were then computed with group as a factor and the same covariates.

Since no pretest was possible on an instructional selection, reading and IQ scores were again used as covariates in the analysis of written recall results of a familiar selection. Univariate analysis results (see Table 1) revealed that the experimental group significantly outperformed the control group on the total number of text structure categories recalled on an instructional selection (familiar to the treatment group, $F\ (1.56)=51.89$, $p<.001$). This result replicates previous findings (Gordon, 1980) and provides further support for the hypothesis that children who have been trained in story schema awareness will utilize it as a metatextual aid in information retrieval.

To test the remaining hypotheses, a univariate analysis of covariance of posttest measures was conducted using scores from corresponding pretests. The second hypothesis was that story structure awareness was a transferable framework that would facilitate comprehension of a similarly organized but unfamiliar narrative. The results (see Table 1) supported this hypothesis in that the experimental group recalled significantly more text structure categories than did control in the written recall of a new selection ($F\ (1,56)=39.17$, $p<.001$).

The third hypothesis postulated was that story schema awareness would facilitate literal and inferential comprehension of an unfamiliar selection as demonstrated by ability to answer literal and inferential wh- probes keyed to specific text structure categories. Results (see Table 1) showed that while both groups had their background knowledge (content schemata) on the topic developed immediately prior to testing, the group with training in story structure awareness answered more questions correctly than did the control group ($F\ (1,56)=15.89$, $p<.001$).

Finally, it was hypothesized that knowledge of story schema would serve as a scaffolding for independently generating information under specific text structure categories. The hypothesis was supported by the fact that the story schema group significantly outperformed the control group on the total number of text structure categories produced in the creative writing task ($F\ (1,56)=4.18$, $p<.05$) (see Table 1).

TABLE 1
Means and Standard Deviations for all Posttest Measures

Dependent Measures	Experimental	Control
	Mean Sd	Mean Sd
Reading Tasks		
Written Recall Familiar Selection (final Instructional Selection) Total Number of Text Structure Categories	9.8 (1.6)	6.0 (2.4)
Written Recall Unfamiliar Selection Total Number of Text Structure Categories	8.1 (2.0)	4.8 (2.4)
Wh- Comprehension Questions on Unfamiliar Selection Total Overall Comprehension	6.7 (1.8)	4.9 (2.0)
Writing Task		
Creative Writing Production Total Number of Text Structure Categories	11.4 (4.2)	9.1 (4.5)

DISCUSSION

The findings of this study provide support for the notion that children who have been trained in story schema awareness employ the structure not only for recall of familiar selections but also for comprehension and recall of new and unfamiliar stories. Story schema appears to be a transferable framework which, once internalized as a metatextual aid, is used for comprehension, storage, and retrieval of information read. Further, instruction in story schema provides a useful schematic framework for children's independent production of stories.

Lack of transfer effects in Gordon's (1980) findings suggested that treatment effects can be task specific and the effectiveness of strategies can vary under specific task demands. In the 1980 study, no provision for transfer was made and children were expected to independently transfer learnings to new and more difficult tasks. Present results show that when an instructional method is designed to deliberately enhance and facilitate transfer, children readily apply story schema to related reading and writing tasks. Therefore, teacher-guidance and student practice on a variety of materials in reading and writing appear to be necessary.

Previous findings (Gordon, 1980) had also shown that knowledge of macrostructure did not appear to significantly aid in the comprehension of new selections. That is, there was little support for teaching text structure as a metatextual aid to literal and inferential comprehension. The results of the present study, wherein careful control was exercised on content schemata by developing each group's background knowledge on the topic to the same extent, provide strong evidence that knowledge of text structure facilitates the drawing of textually-explicit, as well as textually-implicit and scriptally-implicit relationships tied to specific text structure categories.

Further analyses of data will reveal which specific text structure categories and question-answer relationships contributed to significantly overall better performance of the experimental group on each of the posttest measures. Delayed posttest data will provide information on the effect of time on knowledge of story schema in reading and writing.

REFERENCES

Bower, G. H. Selective facilitation and interference in retention of prose. *Journal of Educational Psychology*, 1974, *6*, 1–8.

Bower, G. H. Experiments on story understanding and recall. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 511–534.

Brown, A. Recognition, reconstruction and recall of narrative sequences by preoperational children. *Child Development*, 1975, *46*, 156–166.

Brown, A. Semantic integration in children's reconstruction of narrative sequences. *Cognitive Psychology*, 1976, *8*, 247–262.

Chodos, L., & Mosenthal, P. Fourth graders' comprehension of story structures under three recall conditions. In P. D. Pearson & J. Hansen (Eds.), *Reading: Disciplined Inquiry in Process and Practice*, Twenty-seventh Yearbook of the National Reading Conference, Clemson, South Carolina: The National Reading Conference, 1978.

Gordon, C. J. *The effects of instruction in metacomprehension and inferencing on children's comprehension abilities*. Unpublished doctoral dissertation, University of Minnesota, 1980.

Graves, D. H. An examination of the writing processes of seven year old children. *Research in the Teaching of English*, 1975, *9*, 227–241.

Kintsch, W. Memory for prose. In C. N. Cofer (Ed.), *The structure of human memory*. San Francisco, CA: W. H. Freeman and Co., 1975.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111–151.

Pearson, P. David, & Johnson, D. *Teaching reading comprehension*, New York: Holt, Rinehart and Winston, 1978.

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge & J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Stein, N., & Glenn, C. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *Discourse processing: Multidisciplinary perspectives in discourse comprehension*. Hillsdale, N.J.: Ablex, Inc., 1977.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, *9*, 77–110.

van Dijk, T. A. Semantic macro-structures and knowledge frames in discourse comprehension. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

NORA L. HOOVER
University of Florida

YOUNG CHILDREN'S DETECTION OF INCOHERENCE IN NARRATIVES

In a schema theoretic conceptualization of reading, comprehension is viewed as an interactive process of mapping incoming information to preexisting structures. To the extent that incoming data is congruent with these existing cognitive structures or schemata, to that extent are comprehension and recall possible (Rumelhart & Ortony, 1977; Adams & Collins, 1979). The influence of schemata believed to guide the processing of story information has been fairly well researched. Investigations examining this type of cognitive processing have been made possible through the development of story grammars which attempt to specify the underlying cognitive structures constituting a story schema.

Developmental differences in children's ability to deal with disruptions in the logical structure of a story have been reported in several investigations utilizing story grammars (Stein & Glenn, 1977; Stein & Nezsworski, 1977; Mandler, 1978; Poulson, Kintsch, Kintsch, & Premach, 1979). In these studies, older children consistently demonstrated greater competence in processing disorganized story information than younger children. Thus, previous research suggests that children with well developed story schemata will detect when the cohesiveness of a story is disrupted. Children who do not have an expectation of what constitutes a well constructed story, or who do not engage their story schema in the encoding of a story, are less likely to notice disruptions in the cohesiveness of a story.

In the research cited, a number of questions were not investigated: the range of developmental differences within and between age groups in the ability to detect disruptions in the logical coherence of a story, the degree to which this ability is present at the time of initial instruction in reading, and the developmental changes over an academic year in the monitoring of story information for incoherence. The present study examined these questions and assessed the influence of this aspect of story schema functioning on beginning reading.

Two hypotheses were tested. The first predicted that mean score achieved by subjects on a task measuring the ability to detect disruptions in the logical sequence of a story would increase by grade level and from fall to spring within grades. The second predicted that a significant amount of the variance in beginning reading would be attributable to the ability measured by the experimental task.

METHODOLOGY

Subjects

A sample of 112 participants comprised of 24 four-year-old children and 88 elementary school children was studied. The four-year-old subjects attended a Head Start (HS) pre-school program in a rural Appalachian community; the elementary children, 44 in kindergarten (K) and 44 in first grade (1st), attended a public school in the same locale. A stratified sampling procedure was used to select the elementary children, to assure, in so far as was possible, a range of variability in the academic potential of the subjects. In this procedure, the total population at each grade was stratified high, medium, or low according to performance on standardized readiness tests. A representative proportional sample was then randomly selected from each group.

Procedure

Subjects were tested individually. The experminter gave the following directions: "I'm going to tell you the same story in different ways; sometimes it might sound mixed up, sometimes it will sound OK. You must listen carefully to hear if the story is mixed up or if it sounds alright. After each time I read the story, I'd like you to tell me if it was OK or if it sounded scrambled. Do you understand?" The child was asked to repeat the directions to the experimenter.

The experimenter then read the same story six times including three normal versions and three with structural variations. Presentation order of the six stimuli was random. Two stories were used, both fables rewritten by Nezworski, Stein, & Trabasso (1979) to conform to the Stein & Glenn (1975) grammar. They were therefore structurally similar. Story presentation was random: if a subject heard the Fox and Bear story in the fall, the Tiger's Whisker story was used in the spring and vice versa.

In one of the three variations, the consequence was moved to the initial event position; in another the initial event was moved to the consequence position; in the third, the setting was moved to the consequence position. These categories were chosen to deviate because of their salience in comprehension and recall in previous research (Stein and Glenn, 1977; Mandler and Johnson, 1977).

Scoring and Analysis

In assigning points to the subjects' responses, maximum credit (4 points) was given for patterns of responses indicating the ability to detect variations in story structure and a minimum (0 points) to responses indicating an inability to detect a disruption in the cohesiveness of a story, or to responses given at random.

Between group differences were tested using a one way analysis of variance. Where appropriate, *post hoc* multiple comparisons were made using the Tukey HSD ($p<.05$) procedure to determine which means or combinations differed significantly. The developmental change in story schema over the course of the school year was tested using the appropriate procedures for dependent samples, that is, a Student's t-test.

For the first grade subjects, the prediction of reading achievement was tested using an analysis of covariance procedure that calculated the residualized gain scores from two administrations of the *Gates-MacGinitie Reading Tests* while holding the IQ constant, then tested whether or not there was a statistically significant difference in reading achievement between those children who performed well on the experimental tasks in the fall and those who did not. An analysis of variance was used to determine whether order of presentation resulted in systematic differences in performance. It was reasoned that if children heard the story correctly two or three times in a row they might identify the scrambled versions more readily. Similarly, a normal/scrambled pattern over the six exposures might have resulted in higher scores.

RESULTS

The analysis to determine whether order of presentation affected performance indicated that it had no effect. No significant differences were observed between those subjects who heard the normal conditions first, or in an alternating pattern, and those who did not.

The first research hypothesis relating to within and between group differences was, in general, supported by the data. *Post hoc* multiple comparisons revealed differences that were significant at both the fall and spring administration between kindergarten and first grade subjects ($p<.05$).

Within each group examined, some growth in sensitivity to structural deviations was demonstrated over the course of the year. The amount of improvement was significant for both the kindergarten ($p<.02$) and first grade ($p<.006$) subjects. Means and standard deviations by grade level for performance on the task are presented in Table 1.

TABLE 1
Detection of Incoherence in Narratives

Means and Standard Deviations by Grade Level

	HS		K*		1st*	
	Fall	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**
$\bar{X}$	.25	.44	1.05	1.86	1.80	2.60
SD	.19	.25	1.50	1.70	1.68	1.59

*$p<.02$

With regard to reading achievement, results indicated that a substantial amount of the variance in total reading growth, vocabulary plus comprehension, over the year was attributable to the ability measured in this procedure for the first grade subjects ($MS=398.61$, $F=2.24$, $p<.08$).

DISCUSSION

In the present study, it was predicted that at increasing stages of story schema acquisition, children would have proportionately less difficulty making a distinction between normal and structurally deviated stories. This developmental difference was anticipated because of lack of congruence children would experience between incoming data (the scrambled story) and the developing schema. The better defined the schema, the greater the incongruence. In general, this hypothesis was supported by the findings.

However, by the end of the school year, 30% of the first graders did not recognize the structurally deviated stories as being out of order. These findings are similar to Denhiere and LeNye's (1980) who also reported what appeared to be a metacomprehension failure among young children. In their study of comprehension and recall of narratives by children eight and eleven and adults, the older children and adults recalled essentially the same propositions they had independently judged as most important. The younger children did not recall the units they judged to be important, but, instead, recalled what the older subjects judged important. Thus, the eight year olds evidenced developmentally different metacomprehension abilities than the eleven year olds but no comparable difference in schema processing on an unconscious level.

The results of the Denhiere and LeNye (1980) study, and performance by the first graders in this investigation, suggest that many children, even by the end of first grade, have not yet acquired the capacity for certain types of metacomprehension operations on oral discourse. The differential influence of schemata as a function of its acquisition is suggested as a possible contributor to the observed differences in metacomprehension operations across groups.

An unexpected finding was that of no difference in recognition levels for the three conditions. It was hypothesized that the consequence moved condition would be the most readily identified as a scrambled story because it caused the most disruption in the coherence of the stimuli. These results are similar to those of Thorndyke and Yekovich (1980), who found that the recognition of a story proposition following a reading is independent of its importance in the hierarchy of a story. In this task as well, the recognition of a deviated category was independent of its importance. These findings suggest that schema based recognition operations differ from retrieval mechanisms. In retrieval, the level of a proposition in the schema hierarchy largely determines its accessibility; this does not appear to be the case in certain types of recognition operations.

As anticipated, the ability to distinguish between normal and structurally deviated stories did prove predictive of subsequent reading achievement among the first graders studied. This finding suggests that even at the earliest stages of reading acquisition story schema exerts a significant influence on the processor's ability to monitor narrative material for incoherence. In light of these results, it would seem that increased attention to story schema as a processing mechanism by beginning readers would prove beneficial to reading growth.

REFERENCES

Adams, M. J., and Collins, A. A schema-theoretic view of reading. *New Directions in Discourse Processing*, Vol. 2. Roy O. Freedle (ed.). Norwood, N.J.: Ablex Publishing Co., 1979.

Denhiere, G., LeNye, J. F. Comprehension and recall of stories. *Poetics*, 1980, *9*, 23–49.

Kintsch, W. *Memory and cognition*. New York: John Wiley and Sons, 1977.

Mandler, J. M. A code in the node: The use of story schema in retrieval. *Discourse Processes*, 1978, *1*, 14–35.

Mandler, J. M., and Johnson, N. S. Remembrances of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111–151.

Markman, E. M. Realizing that you don't understand: A preliminary investigation. *Child Development*, 1977, *48*, 986–992.

Nezworski, T., Stein, N. L., and Trabasso, T. *Story structure versus content effects on children's recall and evaluation inferences*. Paper presented at the Psychonomic Society Meeting, San Antonio, Texas, 1978.

Poulson, D. Kintsch, E., Kintsch, W., and Premack, D. Children's comprehension and memory for stories. *Journal of Experimental Child Psychology*, 1979, *28*, 279–403.

Rumelhart, D. W., and Ortony, A. The representation of knowledge in memory. In Anderson, Spiro, and Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Stein, N. L., and Glenn, C. G. *An analysis of story comprehension in elementary school children: A test of a schema*. ERIC, 1975, ED 121 474.

Stein, N. L., and Glenn, C. G. *The role of structural variation in children's recall of simple stories*. Paper presented at the meeting of the Society for Research in Child Development, New Orleans, March 1977.

Stein, N. L., and Nezworski, T. *The effects of linguistic markers on children's recall of stories: A developmental study*. Unpublished manuscript, University of Illinois, Urbana-Champaign, 1977.

Stein, N. L., and Nezworski, T. *The effects of organization and instructional set on memory*. (Technical Report No. 68). Urbana, Ill.: University of Illinois, Center for the Study of Reading, January 1978.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, *9*, 77–110.

R. KAY MOSS
JOHN C. STANSELL
Texas A & M University

THE READER-BASED AND LISTENER-BASED ORAL TEXT PROCESSING OF A SIXTEEN-YEAR-OLD BEGINNING READER

In this paper we wish to present a "language story" (Harste, Burke, and Woodward, 1981), an account of an extended, evolving language situation that provides insight about written language use and instruction. This particular story illustrates 1) the importance of learners' language strengths, 2) the importance of learners' perceptions about the nature of assigned task, and 3) the need to examine reading instruction as a possible contributor to delayed literacy. The story involves two persons who participated directly in this language situation: James, a Black, sixteen-year-old beginning reader; and Kay, James' tutor for a period of weeks and the senior author of this paper.

James was introduced to us through the probation office of the local police department. His probation officer described him as "illiterate". His teachers described him as a "trouble maker". We intend to describe him as a language user.

On the first day of tutoring, Kay took James for a ride through his downtown neighborhood and asked him to read anything that he could. When we encountered a "one-way" traffic sign with an arrow pointing to the right, James read "Turn right". When he encountered "Bryan Public Library" over the entrance to that building, he read "Bryan Library". He read "good rich" from the Goodyear logo on the front of the tire store. From these and other interactions with print, we became aware that James was indeed not illiterate when dealing with such environmental print.

James' probation officer had informed us that he "can't even write his name" and his former English teacher stated "he can't read a lick." Literacy, however, is context specific (Harste, 1979); while James may have been unable to deal successfully with certain kinds of print in some situations, he was at least somewhat successful as a reader in other instances. Even though James was not as successful a reader as we have come to expect a young man his age to be, it is important to recognize that his varying success in different situations is a characteristic that he shares with us, and with all readers.

JAMES' TWO STORIES

After returning from their ride through the neighborhood, Kay turned on a tape recorder and asked James to tell a story. The following is an excerpt from that story which he titled "The Basketball Kid":

> This boy he's 15 years old. He always wanted to play basketball. So he went to a Boy's Club and didn't nobody recognize him. And so everybody come up and ask "What's his name?" He said his name was Wayne. They told him their names. They asked did he want to play basketball? He said he didn't play with chumps.

The story goes on to tell of the misadventures of Wayne, the protagonist of a very imaginative, detailed story featuring dialogue, a well-developed plot, good characterization, and a brilliant climax in which Wayne's basketball career, seriously threatened when he gets into trouble with the law, is dramatically salvaged by his coach.

The week after James told "The Basketball Kid", Kay arrived with a typed transcription of his story. This story was used as the reading material for the next several sessions: he illustrated the story, read along with the tape, and completed a modified cloze procedure where he wrote in predictable words and phrases that had been deleted. Kay's initial interview with him, in which he described his previous reading instruction, leads us to believe that during these sessions James read natural, connected discourse for the first time in his life.

Finally, when James was ready for new material, he taped another story. While Kay settled back with the tape recorder for another creative narrative, James began "Paperboy *vs.* Mailman":

> This boy . . . was throwing . . . newspaper . . . every . . . see . . . this boy throwed newspaper every day. And . . . the mail . . . the mailman . . . was . . . always tries to beat him . . . to the box. The boy always rode a bike. The mailman had to drive a jeep. And the mailman had to stop for the stop . . . sign. And the boy . . . alway . . . rode his bike . . . around the curve . . . on the sidewalk. And . . . so . . . they . . . betted . . . he I mean . . . betted . . . he could beat him. To the mailman. And beat the mailman in. And the boy betted he couldn't. So . . . the next day they had . . . a race. See who was going to be there first. So . . . the mailman . . . saw him . . . coming . . . at the stop sign. And the mailman . . . ran . . . the stop sign . . . and . . . the boy . . . was coming . . . around . . . the curve . . . and the mailman . . . look . . . off . . . see if there was any car was coming. And he drove faster . . . to the . . . house. And . . . he didn't see the . . . boy on the curve. Hit him.

ANALYSIS OF THE STORIES

Our immediate reaction was that this story was remarkably different from the first story. The total number of words in James' first story was 882; in the second story the total number of words was 162. The first story lasted nearly eight minutes while the second story lasted only two and a half, even though James spoke much more slowly (see Table 1) and deliberately ("thē", "āround"). An examination of certain linguistic features within the two stories revealed other important differences.

TABLE 1
Length of Stories in Words and Seconds

	"The Basketball Kid"	"Paperboy *vs.* Mailman"
Total words	883	162
Total seconds	475	151

Idea Units

One feature of oral language is that it is not produced in a continuous stream, but in a series of brief spurts. Halliday (1967) called attention to these spurts of language some time ago, giving them the name "information units". More recently Kroll (1977) and Chafe (1980) use the term "idea units", which we adopt here. Three obvious criteria aid in identifying idea units. The most important criterion is intonation: idea units end with a rise in pitch or a fall in pitch. A second factor is pausing: idea units are usually followed with at least a brief pause, and sometimes with an even more significant pause of a second or more. A third factor is that there is a tendency for idea units to consist of a single clause, which may often begin with the word "and" or, less often, with another conjunction such as "but" or "so".

"The Basketball Kid" consisted of 128 idea units whereas the second story, "Paperboy *vs.* Mailman", only totaled 60 (see Table 2). But even more interesting is the average word count per idea unit. In James' spontaneous conversational speech, the mean length of idea units is about 3.5 seconds, or about six words. The idea units in James' first story are similar to those produced in conversation, indicating a rather natural, spontaneous flow of oral story telling. However, in "Paperboy *vs.* Mailman" the mean length in seconds is over one second shorter, illustrating how choppy his phrasing was. Also, the mean length of idea units is words is markedly less: only 2.7 words per idea unit rather than his previous 6.9. Indeed, when a colleague heard this taped story she thought James was *reading* the story, in the halting, hesitating manner of some beginning readers.

TABLE 2
Length of Idea Units in Words and Seconds

	"The Basketball Kid"	**"Paperboy *vs.* Mailman"**
Idea units	128.0	60.0
$\bar{X}$ no. words per idea unit	6.9	2.7
$\bar{X}$ no. seconds per idea unit	3.7	1.1

Centers of Interest

A second feature of oral language that helped us to see differences in these two stories is what is called a center of interest (Chafe, 1980). A center of interest is the expression of an idea or topic that is too long or complex to be fully developed in a single idea unit. Speakers therefore create centers of interest by linking related idea units together through the use of non-final intonation; usually this is a rise in pitch occurring at the ends of idea units. Centers of interest are somewhat analogous to paragraphs in written language in that they generally develop a single topic.

James developed a total of 36 centers of interest in his first story, and a total of 18 in his second (see Table 3). "Basketball Kid's" centers of interest were also developed fully: he didn't spare words in reporting details of the characters or plot. However, his second story averages only nine words per center of interest; even a brief inspection of the transcript will reveal that it contains only the bare threads of a story line, with no well-developed characters.

TABLE 3
Length of Centers of Interest in Words and Seconds

	"The Basketball Kid"	"Paperboy *vs.* Mailman"
Centers of interest	36.0	18.0
$\bar{X}$ no. words per center of interest	24.5	9.9
$\bar{X}$ no. idea units per center of interest	3.6	3.3

Story Grammars

A final look at the differences between James' two stories, "Basketball Kid" and "Paperboy *vs.* Mailman", involved an analysis of story grammar (Mandler & Johnson, 1977). In this analysis, each story was parsed into episodes according to when the protagonist's goal changed; the number and characteristic of episodes in the two stories were then compared.

This analysis revealed that "Basketball Kid" is structured in a very complex manner (see Figure 1). Two related narratives were actually generated, the first of which introduced the second. The first narrative contained five episodes and the second narrative contained eleven. The episodes are ending-embedded: the story is organized so that the Ending of each Episode is in fact all of the remaining Episodes. The result was a cohesive story with a relatively sophisticated, well-developed plot.

However, in the "Mailman" story, James produced a narrative that is not as richly and fully delineated (see Figure 2). The largest segment of the story was devoted to the establishment of the setting. The remainder of the story was a single episode with a few idea units devoted to the Beginning Event, the Attempt (effort to achieve a goal), the Outcome (the success or failure of the Attempt), and a very brief Ending. The result was a story with an impoverished plot structure and, in general, little to offer an audience.

LISTENER-BASED AND READER-BASED STORIES

Why did James create such different stories? Why did he ignore his considerable strengths as a language-user and story-teller in his second story? We believe he did so because he was aware that his second story would be transposed into print and used for reading instruction.

All speakers must somehow orchestrate a complex array of linguistic and paralinguistic features in order to accomplish the purpose of their discourse. The precise nature of this orchestration is influenced heavily by the setting and situation in which it occurs. When James told his first story, he did not know that Kay intended to transcribe it and ask him to read it. But after several sessions of instruction built around the transcript of that story, there is little doubt that James realized Kay's plans for his second story.

Essentially, the first story James told was a *listener-based story*. Its linguistic and paralinguistic characteristics meet the needs and expectations of a listener; in short, it has the sound, feel, and texture of an oral-language story. As such, it is an oral-language counterpart of reader-based stories (Flower, 1979) in written language.

FIGURE 1. Idea Units in Story Grammar Categories in "The Basketball Kid"

Narrative 1	Idea Unit No.
Setting:	1–2
Episode 1	
Beginning:	0
Simple Reaction:	0
Goal:	3
Attempt:	4
Outcome:	5
Ending:	All that follows
Episode 2	
Beginning:	6
Complex Reaction:	0
Attempt:	7
Outcome:	8
Ending:	All that follows
Episode 3	
Beginning:	9
Complex Reaction:	0
Attempt:	10
Outcome:	11
Ending:	All that follows
Episode 4	
Beginning:	12–16
Simple Reaction:	0
Goal:	17–19
Attempt:	20–22
Outcome:	23–36
Ending:	All that follows
Episode 5	
Beginning:	37–38
Complex Reaction:	0
Attempt:	39
Outcome:	40–44
Ending:	45

Narrative 2	Idea Unit No.
Setting:	46–48
Episode 1	
Beginning:	49
Simple Reaction:	0
Goal:	50
Attempt:	0
Outcome:	51
Attempt:	52
Outcome:	53–56
Ending:	All that follows
Episode 2	
Beginning:	57
Complex Reaction:	0
Attempt:	58–59
Outcome:	60
Ending:	All that follows
Episode 3	
Beginning:	61
Simple Reaction:	0
Goal:	62–64
Attempt:	65–68
Outcome:	69–70
Ending:	All that follows
Episode 4	
Beginning:	71–72
Complex Reaction:	0
Attempt:	73
Outcome:	74
Attempt:	75
Outcome:	76–77
Attempt:	78
Outcome:	79
Attempt:	80
Outcome:	81–85
Ending:	All that follows
Episode 5	
Beginning:	86–87
Simple Reaction:	0
Goal:	88a
Attempt:	88b
Outcome:	89a
Outcome:	89b
Ending:	All that follows

	Idea Unit No.
Episode 6	
Beginning:	90–97
Complex Reaction:	0
Attempt:	98–99
Outcome:	100–102
Ending:	All that follows
Episode 7	
Beginning:	103
Complex Reaction:	0
Attempt:	104
Outcome:	105–106
Ending:	All that follows
Episode 8	
Beginning:	107–108
Complex Reaction:	109
Goal:	110
Attempt:	111
Outcome:	112–113a
Ending:	All that follows
Episode 9	
Beginning:	113b
Complex Reaction:	0
Attempt:	114
Outcome:	115–116
Ending:	All that follows
Episode 10	
Beginning:	117
Complex Reaction:	0
Attempt:	118–119
Outcome:	120–121
Ending:	All that follows
Episode 11	
Beginning:	122–125
Complex Reaction:	0
Attempt:	126
Outcome:	127
Ending:	128–130

FIGURE 2. Idea Units in Story Grammar Categories in "Mailman *vs.* Paperboy"

	Idea Unit No.
Setting:	1–27
Episode	
Beginning:	28–35
Complex Reaction:	0
Attempt:	36–39
Outcome:	40–59
Ending:	60

His second story, however, is in large measure the product of James' realization that it would be transcribed and used for instruction. As such, it represents his attempt to orchestrate language according to the demands of a different situation. A situation which called for a predominance of the features of written stories, as he had experienced them, and which would soon demand that he read this story. It is thus a *reader-based story*, reflecting James' limited prior experience with stories in print and his careful choices of just those grammatical features and lexical alternatives that he felt he could handle as a reader.

Though "Paperboy *vs.* Mailman" may be a rudimentary first attempt at producing a written-language story, it must be recognized that this story demonstrates James' awareness of and sensitivity to the differential characteristics of speech and print. And though he had previously written no more than his name on school papers and legal forms, he has shown us an understanding of the notion of audience in written language as well as in oral language, and of the fact that the expectations of readers are different from those of listeners.

DISCUSSION AND IMPLICATIONS

While we are delighted at the sophistication that James shows in these two stories, we must also be prepared to acknowledge that the two stories are quite different in quality. We must face the reality that "Mailman *vs.* Paperboy" is most probably the unfortunate result of James' written language instruction, which focused his attention on accurate performance in reading and writing and away from meaning.

In interviews with Kay prior to the taping of his first story, James revealed a preoccupation with exactness in reading and with knowing words as crucial to the success of a reading event. He was most reluctant to write during tutoring sessions out of fear that what he produced would be wrong. In contrast to so many young children who freely use functional invented spelling prior to instruction, James would almost rather avoid the whole process than risk being wrong. To him, good reading and good writing both seem to be a matter of being "right", and being right the first time.

Written language instruction that leads learners to develop such notions also tends to offer little in the way of exciting and well-developed stories to be read, and equally little in the way of opportunities for composing real, rhetoricallybased stories. In fact, "Mailman *vs.* Paperboy" resembles in some important ways the pseudostories that appear in too many basals for the purpose of practicing recently-taught word attack skills in a context that can best be described as impoverished.

As we have indicated, "Mailman *vs.* Paperboy" is structurally simple and underdeveloped. There is no dialogue in this story, and we learn little about its characters. They have no names, and no real impact upon the listener. We do not know whether the paperboy was killed or injured, or what sort of effect being hit by the jeep may have had upon him. We do not know if the mailman summoned help, or if he saved the boy's life through the heroic application of first aid, or if he tried to comfort the boy somehow until help arrived, or if he callously left the scene of the accident. While we can easily adopt an aesthetic stance (Rosenblatt, 1978) toward "The Basketball Kid" and share in the protagonist's frustrations and triumphs, it is practically impossible to adopt such a stance toward "Mailman *vs.* Paperboy"; there is nothing in the story to generate feelings within a listener because there is no real human experience to live through.

Given the characteristics of "Mailman *vs.* Paperboy", it may be more appropriate to call it a *reading-instruction-based story*. To the extent that stories of this sort characterized James' reading instruction, it is not too difficult to imagine that they may have rendered him somewhat less than eager to be a reader. As we indicated earlier, James at age 16 was literate only in a narrow range of print settings. Although there is no doubt that many factors are involved in the development of young adults who read no better than James, we would contend that research should seriously investigate reading instruction as a contributing factor.

There is one other important cue that researchers can take from James, which lies in the fact that his second story was neither intended nor predicted to be the kind of story it was. In fact, the reverse was true; we both intended for him to produce another story of the quality of "The Basketball Kid", and would have predicted that he would do so. But when James sat down to tell "Mailman *vs.* Paperboy", the situation as he perceived it was different from the one in which he told "The Basketball Kid"; Kay had clearly become a teacher, and their interaction was now for the purpose of instruction in reading. And it is crucial to note that although the instruction that Kay provided was rooted in real language and focused upon meaning, James' previous experience with reading instruction was such that this difference was not important to him or perhaps not even noted by him. The total situation as he perceived it seemed to call for a story like "Mailman".

As researchers, we must not expect that subjects' performance is limited or affected only by some index of their ability to perform an assigned context-free task. Nor is it enough to be cognizant of the nature of the situation from our own perspective; indeed, that is what led us to our faulty prediction about the kind of story James would tell the second time around. We must instead be prepared to gain an awareness of the situation *as the subject perceives it*. Only then can we begin to develop a better understanding of the sociopsycholinguistic process that reading is.

REFERENCES

Chafe, W. L. (Ed.) *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Vol. III in the Series: Advances in Discourse Processes; Roy O. Freedle (Ed.). Norwood, New Jersey: Ablex Publishing Corporation, 1980.

Flower, L. Writer-based prose: A cognitive basis for problems in writing. *College English*, 1979, *41*, 19–37.

Halliday, M. A. K. Notes on transitivity and theme in English. Part 2. *Journal of Linguistics*, 1967, *3*, 199–244.

Harste, J. C. Reading and reality: a sociopsycholinguistic view of reading and reading instruction. In B. Farr and D. Strickler (Eds.), *Reading comprehension: Handbook*. Bloomington, Indiana: Indiana University Language Education Departments, 1979.

Harste, J. C., Burke, C. L., and Woodward, V. A. Children's language and world: Initial encounters with print. In J. Langer and M. Smith-Burke (Eds.), *Reader meets author: Bridging the gap*. Newark, Delaware: International Reading Association, 1981.

Kroll, B. Combining ideas in written and spoken English: A look at subordination and coordination. In E. O. Keenan and T. L. Bennett (Eds.), Discourse across time and space. *Southern California Occasional Papers in Linguistics* No. 5. Los Angeles: Department of Linguistics, University of Southern California, 1977.

Mandler, J., and Johnson, N. Remembrance of things parsed: story structure and recall. *Cognitive Psychology*, 1977, *9*, 111–51.

Rosenblatt, L. *The reader, the text, the poem*. Carbondale, Illinois: Southern Illinois University Press, 1978.

DIXIE LEE SPIEGEL
JILL FITZGERALD
University of North Carolina-Chapel Hill

CONFLICT AND CONFLICT STRUCTURES IN STORIES TOLD BY CHILDREN

In the last few years many reading researchers have investigated children's and adults' knowledge about stories. Several investigators have suggested that individuals use knowledge about story structures to understand and remember stories (e.g., Mandler & Johnson, 1977; Rumelhart, 1978; Stein & Glenn, 1977; Thorndyke, 1977). Recently, aspects of stories other than story structure have been examined (e.g., Brewer & Lichtenstien, in press; Morgan & Sellner, 1981). Among these are conflict and conflict resolution, which have been suggested as contributing both to the complexity and interest of a story (Steinberg & Bruce, 1980).

In examining higher-level features in stories written for children, Steinberg and Bruce (1980) have defined conflict as occurring when story characters cannot achieve one or more of their desired goals. They propose that conflict may be an essential ingredient of stories in general and that the types of conflict in a story affect the reader's enjoyment of and understanding of the story.

Steinberg and Bruce (1980, pp. 10–11) have identified three kinds of conflict: 1) Environmental, in which attainment of a character's goal is hampered by nature, society, or fate; 2) Interpersonal, in which two or more characters have incompatible goals; and 3) Internal, in which a single character has two or more incompatible goals.

Two other features that Steinberg and Bruce analyzed are response to conflict and conflict resolution. A response to conflict is defined as "a verbalization, a thought, or an action that a character makes subsequent to and related to the conflict" (Steinberg & Bruce, 1980, p. 11). A conflict resolution is defined as "a working out of the conflict or an end to the original conflict" (Steinberg & Bruce, 1980, p. 11) and five types of conflict resolution are described: the goal is achieved, partially achieved, willingly forsaken, changed, or not achieved and the failure is accepted.

Steinberg and Bruce examined conflict, response to conflict, and conflict resolution in stories written for children in trade books, basal readers, and other educational texts. The present study extends this research to examine stories told by children. Specifically, the study had three purposes: 1) to determine if children perceive conflict and conflict resolution to be essential characteristics of stories; 2) to determine what kinds of conflict and conflict resolution are given by children in their stories; and 3) to determine if a relationship exists between presence of conflict in stories told by children and reading achievement and between completeness of conflict structure in stories told by children and reading achievement. It was hypothesized that children's story productions would reflect their knowledge about stories (cf., Applebee, 1978; Botvin & Sutton-Smith, 1977; Stein & Glenn, 1977). Thus, if they considered characteristics of conflict to be essential characteristics of stories, they would include them in their story tellings.

METHOD

Subjects

The subjects were 50 fourth grade students from one suburban/rural elementary school. These subjects were identified as the 50 lowest ranking readers from a pool of 100. Reading achievement level was determined by performance on the *California Achievement Test* (CTB/McGraw-Hill staff, 1978) and teacher judgment. The grade level scores for the 50 subjects had a range of 2.2 to 4.0, with a mean of 3.5. There were 21 males and 29 females.

Procedures

Each subject met individually with an examiner and silently read a story setting, such as "Once there was a candy store owner named Charles." The child was then asked to finish the story orally, making it into the best story s/he could. This procedure was repeated for a second setting. The subjects' responses were tape recorded and transcribed. The order of presentation of the settings was counterbalanced.

Responses were analyzed for six factors: 1) presence of conflict, 2) presence of responses to conflict, 3) presence of conflict resolution, 4) type of each conflict present, 5) type of resolution, and 6) completeness of conflict structure. Interrater reliabilities between two judges for these various scoring judgments were .88, .83, 1.00, .85, .85, and 1.00, respectively.

RESULTS

Where possible, statistical tests (Sign Test [Hull & Nie, 1981]) of differences in findings between stories were performed, generally revealing no significant differences. One exception will be discussed in the section concerning types of conflict and resolution. The results reported in all other sections are based on data collapsed across stories.

Presence of Conflict, Response, and Resolution

The first purpose of the study was to determine if children perceive conflict, response to conflict, and conflict resolution to be essential characteristics of stories. The fourth grade readers included conflict in 72% of the stories they told. When conflict was present in a story, the mean number of conflicts per story was 2.41. Therefore, in the majority of cases children did perceive conflict to be an essential characteristic of stories.

When the subjects did include conflict in their stories, they were also likely to include full responses to the conflict. That is, if they included Interpersonal or Internal conflicts, they were likely to give responses for both goals involved in the conflict; for the Environmental conflicts, only one goal was involved and therefore only one response was needed. Eighty-one percent of the Environmental conflicts had responses, 90% of the Interpersonal conflicts had full responses and 100% of the Internal conflicts had full responses. There were significant differences (χ^2 [2] = 2.88, $p<.001$) in percentages of responses given across conflict types.

When conflict was present, the children tended to provide resolution to the conflict. Ninety-three percent of the Environmental conflicts, 93% of the Interpersonal conflicts and 100% of the Internal conflicts had resolutions. There were significant differences for presence of resolution across conflict types (χ^2 [2] = 24.66, $p<.001$) within each story.

Additionally, when fourth grade readers told stories that included conflict, they also tended to give both responses and resolutions. For 92% of all of the conflicts both responses and resolutions were given.

Types of Conflict and Resolution

The second purpose of the study was to identify the types of conflict and resolution that children give when they tell stories. Table 1 summarizes the data for types of conflict given and shows that the type of conflict given is dependent on the story setting that was provided. For Story 1, the most frequently given type of conflict was Interpersonal (60%), followed by Environmental (35%), with Internal conflicts accounting for only a small percentage (4%) of the conflicts. For Story 2, the percentage of Internal conflicts is similarly small (3%). However, the pattern for percentages of Environmental and Interpersonal conflict differs from the pattern for Story 1, with Environmental and Interpersonal conflicts each accounting for 49% of the conflicts for Story 2. Significant differences in percentages of conflict types were found within Story 1 (χ^2 [2] = 17.56, $p<.001$), within Story 2 (χ^2 [2] = 15.06, $p<.001$), and for both stories combined (χ^2 [2] = 28.40, $p<.001$).

When resolutions were given, there were significant differences in the types of resolutions given to Environmental conflicts (χ^2 [2] = 7.21, $p<.028$) and for Interpersonal and Internal conflicts combined (χ^2 [5] = 33.47, $p<.001$). The Environmental conflicts were analyzed separately because resolutions to Environmental conflicts involved only one goal, whereas resolutions to Interpersonal and Internal conflicts involved a nexus for two goals, such as Achieve-Fail. The two most prevalent resolution types for Environmental conflict were Achieve and Fail, accounting together for 82% of the resolutions. For 10% of the Environmental conflicts, a No Resolution type was given; that is, the child made no attempt to supply a resolution.

The most prevalent resolution type for Interpersonal and Internal conflicts was the Achieve-Fail nexus, accounting for 75% of the resolutions. In six percent of the stories, the child did not attempt to supply a resolution.

Relationship with Reading Achievement

The third purpose of the study was to determine if a relationship existed between presence of conflict and reading achievement and between completeness of conflict structure and reading achievement. The most fully developed conflict in a child's story was scored for completeness using a weighted scale from 0 to 4. A score of 0 indicated no conflict, a 1 indicated conflict but *no* response or resolution, a 2 indicated conflict and at least one response, a 3 indicated conflict and resolution, and a 4 indicated the presence of conflict, response, and resolution.

Better readers, as identified by performance on the Total Reading section of the *California Achievement Test* (CTB/McGraw Hill staff, 1978), tended to include conflict in their stories significantly more often than did poorer readers (r_{xy} (49) = .36, $p<.005$). In addition, better readers tended to produce more complete conflict structures than did poorer readers (r_{xy} (49) = .36, $p<.005$). Since the range of reading ability in the sample was restricted to average and

TABLE 1
Percentages of Conflict Types

Story	Conflict Type Environ.	 Interp.	 Internal
1	35	60	4
2	49	49	3
1 & 2	42	55	3

below average readers, it is possible that if above average readers were included in the study, greater variance would have been introduced, possibly resulting in higher correlations.

CONCLUSIONS AND DISCUSSION

Several conclusions may be drawn from the data. First, average and below average fourth grade readers do tend to perceive conflict to be essential to "storiness." Second, when students do provide conflict, they also tend to provide responses and resolutions to the conflict. That is, they seem to have a well-developed sense of all that a conflict entails.

A third conclusion is that these fourth grade readers had relatively unsophisticated ideas about conflict resolutions. They did tend to include resolutions, but most of their resolutions were of the simple Achieve-Fail variety rather than of the more sophisticated types of resolutions found by Steinberg and Bruce (1980) in stories written for children.

The final conclusions are related to reading achievement. Inclusion of conflict and the degree of conflict structure development are each related to reading achievement.

One interesting sidelight is that the trends and patterns in conflict types shown in Table 1 to some extent parallel the findings of Steinberg and Bruce when they examined story materials written for children. They had judges identify the type of the two most important conflicts in each story and found 34% of the conflicts to be Environmental, 59% to be Interpersonal, and 7% to be Internal. These findings are similar to the results of the present study for Story 1 and for Stories 1 and 2 together. These two sets of data together suggest that the kinds of stories children read or listen to may influence their perception of what a story is all about.

A last point concerns the issue of the child's concept of story. The results of several previous studies (e.g., Applebee, 1978; Glenn & Stein, in press; Mandler and Johnson, 1977; Stein & Glenn 1977; Thorndyke, 1977) have indicated that structural features are part of many individuals' concept of an idealized story. The data from the present study indicate that the features of conflict, response to conflict, and conflict resolution may also be part of children's concepts of an idealized story. Integration of these findings and investigation of other factors, such as the affective factors of suspense and surprise (Brewer and Lichtenstein, in press), may lead to a more complete picture of what constitutes a story.

REFERENCES

Applebee, A. N. *Child's concept of story: Ages 2–17*. Chicago: University of Chicago Press, 1978.

Botvin, G. J., & Sutton-Smith, B. The development of structural complexity in children's fantasy narratives. *Developmental Psychology*, 1977, *13*, 377–388.

Brewer, W. F., & Lichtenstein, E. H. Event schemas, story schemas, and story grammars. In A. Baddeley and J. Long (Eds.), *Attention and performance IX*. Hillsdale, N.J.: Lawrence Erlbaum, in press.

CTB/McGraw-Hill Staff. *California Achievement Tests*. New York: McGraw-Hill, 1978.

Glenn, C. B., & Stein, N. L. *Syntactic structures and real world themes in stories generated by children* (Technical Report). Champaign, Ill.: University of Illinois, Center for the Study of Reading, in press.

Hull, C. H., & Nie, N. H. (Eds.), *SPSS update 7–9: New procedures and facilities for releases 7–9*. New York: McGraw-Hill, 1981.

Mandler, J. M. A code in the node: The use of a story schema in retrieval. *Discourse Processes*, 1978, *1*, 14–35.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, 9, 111–151.

Morgan, J. L., & Sellner, M. B. Discourse and linguistic theory. In R. J. Spiro, B. C. Bruce and W. F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence and education*. Hillsdale, N.J.: Lawrence Erlbaum, 1981.

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge and J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Stein, N. L. *The effects of increasing temporal disorganization on children's recall of stories*. Paper presented at the Psychonomic Society Meetings, St. Louis, 1976.

Stein, N. L., & Glenn, C. G. *A developmental study of children's construction of stories*. Paper presented at the Society for Research in Child Development meeting, New Orleans, 1977.

Steinberg, C., & Bruce, B. *Higher-level features in children's stories: Rhetorical structure and conflict* (Reading Education Report No. 18). Champaign, Ill.: University of Illinois, Center for the Study of Reading, 1980.

Thorndyke, P. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, 9, 97–110.

MARGARET A. ATWELL
California State College, San Bernardino

THE AUTHOR AS READER: A COMPARISON OF PROFICIENT AND LESS PROFICIENT COLLEGE WRITERS

This study (Atwell, 1980) examined the thesis that we do a lot of reading in the process of writing any piece of text. Every author knows that writing is seldom if ever, a straight-ahead, linear transliteration of fully-formulated thought into at-once complete discourse. Most writing occurs in bits and bursts, a few utterances written between many pauses. Several things seem to occur during these hesitant moments—remembering, planning, reformulating and, certainly, reading. Certain recent studies have implied that a relationship exists between reading and the act of producing text. Some of these studies characterize reading as a subprocess incorporated in the larger activity of writing (Pianko, 1979, Bridwell, 1979, & Matsuhashi, 1979) while others describe it as an associate process that resides alongside writing (Perl, 1979). In all, these intriguing descriptions lead to this study's attempt to examine directly the role of reading in the composing process.

Composing in language is a synergistic event and in order to reflect this most-important perspective, the findings describe writing to be simultaneously a *product* and a *process* involving not only *interactions* between the two but reader/text *transactions* as well.

While this study only begins to deal with linguistic synergy, it does present a broad picture which, when coupled with other studies detailing particular aspects of writing, suggests ways to address the complexity of visible language.

METHOD

Procedure

Examining the role of reading in writing demands creating an experimental setting in which reading alternately can and cannot occur. To do this, that process was divided via a technique called *blind writing* suggested by Britton and his colleagues (1975). Specifically, ten Traditional undergraduates identified by standardized tests and placement counseling to be above average in writing ability and ten Basic Skills students identified as needing remedial instruction in writing, wrote one personal essay in a timed and videotaped session. During half of this episode, the authors wrote and planned as they normally would in an impromptu situation. During the other half of the episode, however, each author used a pen without ink and wrote on paper textured to allow no visible trace for the authors to read. Carbon paper and a second sheet of paper were affixed to the top sheet so that a visible copy of the essay could be used in the subsequent analyses. Following the twenty-minute writing episode, each author read his complete product and retrospected about the entire writing episode and the strategies he employed.

All comments were categorized as to level of focus toward discourse and served as an initial indicator of the differences between proficient and less proficient writers.

RESULTS

One of the most salient features that characterized the writers was the manner in which they approached their writing. Good and poorer authors did not view text in the same way and evidently planned for writing differently. Half of the comments made by the proficient students focused on their degree of success in stating a clear message and on the overall structure of their discourse. Less able students, in contrast, focused on spelling, punctuation and word choice. As they examined their papers these writers studied their words, often times looking for mistakes. This characteristic proved most significant, recurred throughout the analyses, and supports other descriptions of the less able writer (Shaughnessy, 1977, Perl, 1979).

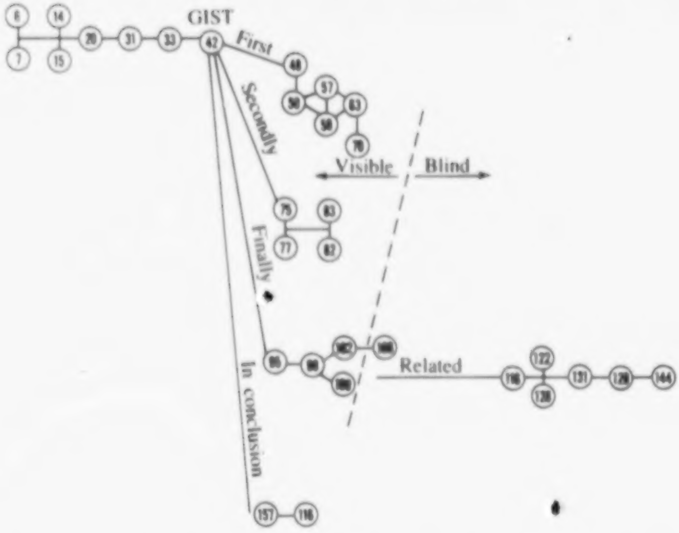
The Product. The analysis of the written products was limited to the semantic structures of the texts and focused on two levels: the microstructure, or the small units of meaning used to express the messages and the macrostructures, or the coherence networks created to tie together the text bases. This included a *local* level of coherence, in the manner in which one constituent connected to the next and a *global* coherence, in how episodes connected to an overriding gist statement.

The first phase of this analysis entailed propositionalizing (Kintsch, 1974, Turner & Greene, 1977, van Dijk, 1977) each text to make comparisons. Although on initial reading the texts seemed to judges to hold differences in meaning and structure, this analysis indicated no significant differences between the make-up of the text bases in the visible and blind writing for either group.

The writing that was done without reading was built out of the same units as that which was written visibly and indicated that, if meaning of an evolving text is affected by the writer's ability to read what he is saying, it does not involve isolated bits.

Any difference seemed then to be related to the manner in which the individual units of meaning were connected into structured discourse. This became the second phase of the product analysis and was approached by mapping the microstructure to represent spacially what had been done verbally in the essays. The maps revealed vast differences between the connectedness of the texts. Two examples, representative of the groups, are included here. While there were clear differences in the degree of connectedness that was achieved by writers of varying ability, these essays also gave some indication that all authors can be susceptible to deviation from plan while writing without the aid of reading.

Writer A had briefly outlined before she began writing and referred to her outline as she wrote, a practice that she later reported she always uses. She further described a formula, sometimes called the "Five Paragraph Theme", on which she commonly relies in impromptu situations. Her outline, essay and map reflect this structuring. The first paragraph introduced the topic and stated a main idea or "gist", while each of the next three paragraphs explicated the topic. Through the use of macroconnectors ("first"; "finally") this writer clearly tied each episode to the gist and divided each unit from the others. A final paragraph tied the episodes together in a conclusion that reiterated the gist. The result was a tightly constructed if somewhat predictable essay. One notable exception to her plan occurred during the third episode and immediately after switching to the blind condition. Here the author deviated from her outline to include an anecdote, the only narrative section within her essay. The map de-


GIST
First
Secondly
Finally
In conclusion
Visible
Blind
Related
6
7
14
15
20
31
33
42
75
77
83
82
131
144
157
116

Coherence Map of a Well-Formed and Highly Coherent Essay

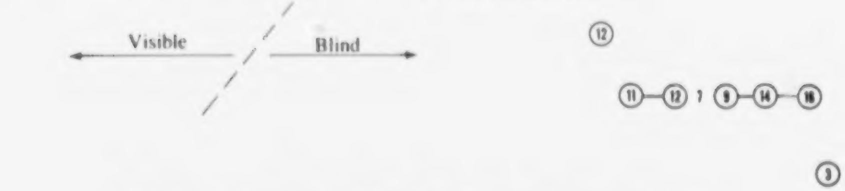

First
9
15
19
27
29
35
36
42
47
48
62
6
13
11
6
5
8
12
11
12
9
14
16
3
Visible
Blind

FIGURE 1. Map of an Essay with Moderate Local Coherence and No Global Coherence

picts this as a linear string, unique in her structuring. While the author clearly controlled her deviation from plan it is interesting to note that it occurred only once and in the absence of the visible record.

His retrospective comments indicate Author B had no evident plan in mind and he appeared to build his essay piece by piece—a task he found to be difficult without the assistance of reading. This essay's map revealed a high degree of local coherence in the visible condition but almost none in the blind and little if any global structure. This author began his essay without stating a superordinate concept that could have acted as an anchor upon which to tie the episodic parts. Failing this, his essay achieved little overall global structure, began confusingly and became increasingly ambiguous as the writer moved through the blind condition.

TABLE 1
Frequency of Occurrence of Process Behaviors Ranked for Condition and Ability

Proficient				Less Proficient			
	V		B		V		B
1. Short pause	(133)	1. SP	(158)	1. SP	(74)	1. SP	(217)
2. Read/visible	(51)	2. WWS	(28)	2. R/V	(25)	2. WWS	(18)
3. Within word stop	(25)	3. LP	(19)	3. MULT	(19)	3. CO	(16)
4. Multiple	(19)	4. CO	(14)	4. WWS	(18)	4. LP	(11)
5. Crossout	(10)	5. MULT	(10)	5. CO	(6)	5. VERB	(10)
6. Read/outline	(8)	6. R/V	(8)	6. LP	(5)	6. MULT	(8)
7. Long pause	(7)	7. R/D	(3)	7. LA	(3)	7. R/V	(7)
8. Read directions	(3)	8. R/O	(2)	8. VERB	(1)	8. R/B	(3)
9. Verbalize	(2)	8. INS	(2)	8. FS	(1)	9. LA	(1)
10. Look away	(1)	10. LA	(1)	13. INS	(0)	9. R/D	(1)
10. Insert	(1)	13. VERB	(0)	13. R/D	(0)	13. INS	(0)
13. False start	(0)	13. FS	(0)	13. R/O	(0)	13. FS	(0)
13. Read attempts/blind	(0)	13. R/B	(0)	13. R/B	(0)	13. R/O	(0)
TOTAL: 505	260/51%		245/49%	444	152/34%		292/66%

When averages for local and global coherence were compared, significant differences were found. The proficient writers attained a high degree (.92) of local coherence when they could read, which they maintained in the blind condition (.93). The less able writers produced moderately coherent (.67) text visibly and lost coherence when they could not read (.31), ($t_{19}=5.41$, $p<.01$). Globally, the able writers' texts were far more connected (.67 to .31) than those of their less able peers ($t_{19}=4.17$, $p<.01$).

The Process. The process description characterized the writing act by noting the frequency of thirteen specific behaviors under visible and blind conditions. As in the product analyses, the proficient students emerged stable across conditions while the less able students were not.

Of direct importance to this study was the frequency of reading. Notably, when it was possible, all the writers in the study read. Proficient students were observed to use reading as a writing activity twice as often than did the poorer writers.

The comparative increase in the frequency of process behaviors across conditions for the less able students suggested that, for them, blind writing was drastically less fluent than writing that is accomplished with reading. As noted above, when it was available, these writers had relied on reading and its removal must account for some of their increased hesitation in blind writing. The findings then support the notion that poor writers, like better writers, are readers of their texts. While the poorer writers in this study did not read as often as the more skilled authors, and this is, no doubt, an important difference, it would be too simple to conclude that less proficient writers produced poorer texts or used a less fluent process only because they did not read or because reading was not as important to them. In this study these writers consistently exhibited a word-level focus toward text and gave little evidence of global planning that might reflect a mental scheme for a connected essay. Not having a clear mental plan, they were, indeed, text-bound and needed to read their texts in order to keep the process moving. Combined with their penchant for in-process error-searching, these authors' process seemed to emphasize editing over a short-circuited creative process. In contrast, the better authors, as a group, had evidenced a discourse-level orientation and revealed that they could conjure up several ways to connect discourse. These students, when not allowed to read their writing, could rely on a mental text to keep their composing process recursive and stable.

The Interaction. Neither the writing process nor the written product exist in isolation. The product and the process create each other as the message at hand is evolved. The analyses of interactions that were developed in this study were limited attempts to relate the writing process to syntactic and semantic features of the texts. Each process behavior was categorized as having occurred in one of four syntactic locations (within a clause, within a word, or at terminal or non-terminal juncture). There is some indication from the findings that able students write in larger units than do less able writers. More often the proficient students tended to stop at terminal junctures. This indicates that when these students paused in their process it was often between sentences and this seemed to help them establish a fluency that resulted from writing in fuller syntactic bursts. The poorer writer tended to pause within words, as if considering spelling or lexical choice, far more often than the other group. Given the discourse-level focus of the proficient students and the word-orientation of the other students exhibited throughout the prior analyses, these findings add to the profiles of the groups.

However, it is hazardous to infer that syntactic location can disclose what was going on in a writer's mind as he pauses. An author may be considering a range of semantic, syntactic, lexical or pragmatic factors whenever he pauses and any further interpretation of these findings is not made.

A similar analysis of the interaction between the process and the semantic structure of the texts was also attempted. Semantic structure, however, does not hold the linearity of syntax and attempts to place the behaviors within structural semantic units became problematic. Although this study does not report any description of process/semantic structure interaction, it identifies this as an important area to be explored.

The Transaction. Every text event incurs a transaction between a reader and a writer linked through discourse. The analyses reported thus far were text-based and, in the absence of a reader, were relatively meaningless. If the findings of the analyses describe meaningful differences, then readers should respond to the texts in different ways.

Three outside readers read the twenty texts to select the best and the worst samples. They chose two as exceptionally poor and agreed on one essay as the best. These readers seemed to be particularly sensitive to local and global coherence, preferring texts that reflected tight structure to those that were looser, or that exhibited fluctuation in coherence. The readers also preferred text that had been produced in a fluent manner with global planning by the author. Reader judgment of text quality supported the indices that had been developed in the study and was taken as a verification of them.

CONCLUSIONS

The premise underlying this study was that there is a dynamic interrelationship between reading and writing in the composing process, a notion that was supported by the findings. Not one writer involved in the study had produced his text without, at some point, stopping in the forward flow of his discourse to read. Reading, these data suggest, is intrinsically related to the recursiveness of writing.

Several researchers have identified *recursiveness* as the most salient feature of writing, a quality that allows the writer to cycle forward and back as a message evolves. This study cites reading as one of two sources of recursiveness in text. Reading allows a writer to return in the text to pick up threads of the discourse, recall what has been released from immediate attention, or create new thought. The second source of recursion is a mental plan or scheme for text that acts as a guide in the production of discourse. The result of such a plan is that it immediately delimits somewhat the author's semantic field. A plan may imply an audience, a tone or stance, or a structure that impels the author in logical directions and discourages less predictable outcomes. While reading allows the author to return in the text to monitor or recall, mental plans enable the writer to go forward or back in mind. Full recursiveness is the quality that results from a combination of these two. Together, reading and planning allow a writer the power to use text as he produces text.

In this study then the proficient writers' orientation encouraged them to be more fully recursive in their writing. These writers indicated that they formed superstructures that they used to direct their discourse. As they wrote, they often stopped to read the emergent text, matching it against their plan, modifying the text or the plan if necessary, editing at several levels. When these writers were not allowed to read their texts, they resorted to an increased reliance on their plans in order to keep their writing recursive and stable across conditions.

The less proficient writers often gave little indication of overall planning. Although it may be possible to build discourse from words and small syntactic units, it is a phenomenon that is text-bound and requires the visual processing of reading. These authors, when not allowed to read, lost their primary means of recursiveness in writing and began to string together words that seldom evolved into true texts.

In all, the findings indicate that we can no more understand writing by studying the written text outside the writer, or the process that brought it to be, than we can understand reading by isolating it from text or reader. The lines that have been drawn between reading and writing in the composing process are arbitrary, and to focus on each in isolation can only be a distraction in understanding recursiveness (the language process fundamental in the evolution) of any text.

REFERENCES

Atwell, M. *The evolution of text: The interrelationships of reading and writing in the composing process*. Unpublished doctoral dissertation, Indiana University, 1980.

Bridwell, L. *Revising strategies in twelfth grade students' transactional writing*. Unpublished dissertation at the University of Georgia, 1979.

Britton, J., Burgess, L., Marten, N., McLeod, A., & Rosen, H. *The development of writing abilities (11–18)*. London: Macmillan Educational Ltd., 1975.

Kintsch, W. *The representation of meaning in memory*. New York: John Wiley & Sons, 1974.

Matsuhashi, A. *Producing written discourse: A theory-based description of the temporal characteristics of three discourse types from four competent twelfth grade writers*. Unpublished doctoral dissertation, State University of New York at Buffalo, 1979.

Perl, S. The composing processes of unskilled college writers. *Research in the Teaching of English*, 1979, *13*, pp. 317–336.

Pianko, S. A description of the composing processes of college freshman writers. *Research in the Teaching of English*, 1979, *13*, pp. 5–22.

Shaughnessy, Nina P. *Errors and Expectations*. New York: Oxford University Press, 1977.

Turner, A., & Greene, E. *The construction and use of propositional text base* (Technical Report No. 63). Boulder, Colorado: Institute for the Study of Intellectual Behavior, 1977.

van Dijk, T. Semantic macro-structure and knowledge frames in discourse comprehension. In M. Just and P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

SHEILA G. COHEN
CONNIE A. BRIDGE
PETER N. WINOGRAD
University of Kentucky

SEMANTIC ANALYSIS OF CHILDREN'S WRITING

A substantial body of research involving children's syntactic development already exists. The work of Hunt (1965), O'Donnell, Griffin, and Norris (1967), Loban (1963), Strickland (1962), Endicott (1973), and Golub and Kidder (1974) has provided researchers with a basic understanding of the developmental trends in children's syntactic development. However, little research has yet been done in the area of developmental changes in the semantic ability demonstrated in children's writing.

King and Rentel (1981) have examined one aspect of children's semantic ability as they described children's use of cohesion to convey meaning in written language without the contextual support typical of oral language situations. Using Halliday and Hasan's (1976) system of cohesion analysis, they found that as children matured in their ability to make meanings clear within the text itself, they made increasing use of lexical ties (vocabulary relationships), conjunctions, and endophoric reference, while their use of exophoric reference decreased.

Odell (1977) has identified linguistic cues or features of the surface structure of written language which indicate the use of various intellectual processes by a writer. These linguistic cues, which include focus, contrast, classification, change, physical context, and sequence, have been listed and described by Odell, but he has not suggested a quantifiable method of measuring the presence of these cues in students' written products other than a simple listing of the cues.

The possibility of quantifying these and other semantic cues exists within Turner and Greene's (1977) adaptation of Kintsch's system for representing the semantic content of a text. A propositional analysis of a text reflects the underlying knowledge structure of the author of the text. A proposition, according to Kintsch's system, is comprised of two or more word concepts, the first being a relation, the others are arguments. Depending on the type of relation which exists between the arguments, three types of propositions can be identified: predication, modification, and connection.

Predicate propositions describe states of actions:

Example 1. (RIDE, I, BIKE)
Example 2. (RECEIVE, I HURT)

Modification propositions change, restrict or limit a concept. They include qualifiers, quantifiers, partitives and negatives.

Example 3. (QUALITY OF, BIKE, BIG)
Example 4. (QUALIFY, BIKE, MY)

Connective propositions relate propositions or concepts either explicitly or implicitly. Connectives express conjunction, disjunction, causality, purpose, concession, contrast, condition or circumstance.

Example 5. (TIME: WHEN 2, 1)

The connective propositions should represent the intellectual processes which Odell (1977) has labeled as contrast, change, physical context and sequence.

The modification propositions should represent the process of classification. It was predicted that as children mature there would be a higher percentage of modification and connective propositions in their written products.

Research Questions

The major purpose of this research was to identify and quantify the types of propositions children use in their writing and to determine whether differences exist in the relative frequency with which various types of propositions are used by children as they develop. A secondary purpose of the research was to see if a semantic analysis of children's composition would contribute additional information above and beyond that gained from a syntactic analysis. The research questions were:

1. Is there a significant difference in the relative frequency of various types of propositions found in the narrative compositions of second, fourth, and sixth graders?
2. Is there a relationship between the proportion of modification and connective propositions in the narrative compositions of second, fourth, and sixth graders and various syntactic measures of these compositions?

METHODS

Subjects and Procedure

Writing exercises were conducted in grades two, four, and six at a public elementary school in a middle class neighborhood in Lexington, Kentucky. Students were requested to write a personal narrative in the following manner: Remember the day when you got hurt? Think about what happened on that awful day." After a brief period of discussion students were told: "Now write a story about that time when you got hurt. Begin your story with "I got hurt when . . ."

Of the total number of compositions students produced, forty-six papers were randomly selected for analysis, fifteen from the second grade, sixteen from the fourth, and fifteen from grade six. Propositions for each composition were analyzed to enable the researchers to identify the relative frequency of various types of propositions which children at various grade levels used in their writing. The compositions were also subjected to a T-unit analysis (Hunt, 1965) to provide measures of words per T-unit, number of clauses per T-unit, and number of words per clause.

RESULTS AND DISCUSSION

Question 1

A preliminary analysis using MANOVA (Hull & Nie, 1981) was performed to examine the differences among the three grade levels for four dependent variables: number of words, number of propositions, number of T-units, and number of clauses in the compositions. Because the test of Wilk's lambda produced an *F* that was statistically significant ($p<.001$; $df=8,80$), univariate ANOVAs were computed and the resulting *F* ratios for each dependent measure were examined to determine which measures contributed to the statistically significant MANOVA results. As expected, second graders produced significantly fewer words, $F(2,43)=25.94$, $p<.0001$; propositions, $F(2,43)=17.95$, $p<.0001$; T-units, $F(2,43)=21.49$, $p<.0001$; and clauses $F(2,43)=20.18$, $p<.0001$ than did the fourth and sixth grades. Differences on these measures between fourth and sixth graders did not reach significance.

Because the compositions of the fourth and sixth graders were significantly longer and contained more total propositions and clauses than those of the second graders, the scores were converted to percentages. Of special interest were the percentages of connective and modification propositions which children used at various grade levels. The proportion of connective propositions found in the protocols was .27 for second graders, .21 for fourth graders, and .21 for sixth graders. The proportion of modification propositions found in the protocols was .29 for second graders, .31 for fourth graders, and .36 for sixth graders. To see if these proportions varied significantly, a 3(Grade) by 2(Type of proposition) repeated measures ANOVA was performed. It revealed no significant main effect for Grade, but there was a main effect for Type, $F(1,43)=25.22$, $p<.0001$. More importantly, there was a significantly Grade by Type interaction, $F(2,43)=3.76$, $p<.03$.

Follow-up tests of simple effects revealed significant differences between grade two and six for proportion of modification propositions ($p<.05$), with sixth graders using proportionately more modification propositions than second graders in their compositions. This had been an expected result. However, contrary to predictions, second graders were found to have a significantly higher proportion of connective propositions in their compositions than did either the fourth or sixth graders. Fourth and sixth graders had almost the same proportion of connective propositions in their compositions.

A closer examination of the second graders' compositions revealed that the unexpectedly high frequency of connective propositions was probably due to their heavy reliance upon the stimulus given at the outset of the writing assignment. All of the second graders used the stem "I got hurt when . . . " to begin their writing. Even though the fourth and sixth graders were given the same instructions, many of them chose to begin their narratives with other introductory sentences. Furthermore, some of the second graders used the stimulus clause in all of the sentences they produced. Thus, instead of developing a cohesive narrative about a single instance when they got hurt, they simply listed several instances when they got hurt. Other second graders merely wrote one sentence using the introductory clause. The result of all of these types of responses to the stimulus was a relatively high percentage of connectives, particularly those of circumstance relating to time. Before the results regarding use of connectives in narratives can be generalized, the study needs to be replicated with a stimulus that avoids the use of an introductory clause containing a connective.

An examination of the protocols revealed that in addition to the connectives in the introductory clause, these second graders were using other connective propositions of circumstance (time and location), concession, and purpose in their writing. At the fourth grade level, propositions expressing condition, causality, disjunction, and circumstance (manner) appeared in some students' compositions in addition to the types of connectives used by second graders. Propositions of contrast did not appear at all even in the sixth grade writing samples.

It has been predicted that connective elements other than "and" would appear less frequently in young children's writing, because of previous research in children's comprehension which showed that third graders, especially poor readers, had difficulty comprehending causal and logical relationships expressed by connectives (Bridge, Tierney, and Cera, 1978). Even ninth graders had difficulty supplying connectives in cloze passages designed to assess their understanding of these conjunctive elements (Bridge, Thomas, and Evers, 1980), and less fluent adult readers experience difficulty with some conjunctive relationships (Marshall and Glock, 1978–1979).

On the other hand, Stein and Trabasso (1981) note that the young children in their study did not have difficulty understanding causal relationships. Stein and Trabasso attribute contrary findings in previous studies to the particular context of the materials to be comprehended. Perhaps when children are reading or writing about familiar topics and events in their own lives, as they were in the present study, they are able to comprehend and express connective relationships, but when they are reading about an unfamiliar topic, the lack of prior knowledge may prevent them from comprehending conjunctive relationships. Other factors which may hinder children's comprehension of conjunctive relationships are the discourse type or task difficulty.

It also must be noted that it was possible for the subjects in the present study to express connective relationships without stating cue or signal words. Using Kintsch's system, relationships need not be stated explicitly. More research needs to be carried out in this area to determine which types of connectives can be comprehended and produced by children at various stages of development and what conditions affect the comprehension and production of connectives.

Question 2

To answer question 2 regarding the relationship between the semantic measures of major interest in this study (proportion of connective and modification propositions) and more traditional syntactic measures (words per T-unit, clauses per T-unit, and words per clause), correlations were calculated between the selected semantic and syntactic measures. It was felt that if the correlations were high, it would indicate that a propositional analysis of subjects' compositions might not yield enough additional information to warrant the time it takes to employ text analysis procedures.

The correlations between these semantic and syntactic measures for each grade separately and the combined grades appear in Table 1. When the measures were combined across grades, those which were most highly correlated were percentage of connective propositions and clauses per T-unit. Since a connective must be used when there is more than one clause per T-unit, it is not surprising that these two measures are correlated. However, it is possible to have connectives which tie two propositions together but do not appear in the same clause or even in the same sentence, so that a correlation of .46 is reasonable. In fact, an analysis of the correlations for each grade taken separately indicate that while second and sixth graders used connectives to include more clauses per T-unit, fourth graders did not. At least for the subjects in this sample, there did not appear to be a developmental trend in the tendency to increase the number of clauses per T-unit through the use of connectives.

Another correlation of interest was length of clause and percentage of modification propositions. If subjects had increased clause length by inserting quantifiers and qualifiers within the same clause, then the correlation should have been high and should have increased with age. Based upon previous syntactic analyses of children's writing, it has been shown that length of clause and

TABLE 1
Correlation Between Syntactic and Semantic Measures by Grade and for All Grades Combined

	SYNTACTIC MEASURES		
SEMANTIC MEASURES	Words/T-unit	Clauses/T-unit	Words/Clause
Grade 2			
Percent of Modification Propositions	.27	−.19	.48
Percent of Connective Propositions	.63	.44	.34
Grade 4			
Percent of Modification Propositions	−.11	.27	.27
Percent of Connective Propositions	−.19	.16	.16
Grade 6			
Percent of Modification Propositions	.08	−.04	.19
Percent of Connective Propositions	.30	.59	−.19
All Grades Combined			
Percent of Modification Propositions	.06	−.06	.21
Percent of Connective Propositions	.08	.46	−.13

T-unit increases with age, and that one way to increase clause length is to insert modifiers within the clause. However, the correlation between percentage of modification propositions and words per clause was highest for the second graders and decreased in fourth grade and even further decreased in sixth grade. Thus, the older children must have been using other methods of increasing clause length than insertion of modifiers within the clause.

CONCLUSIONS

It does appear that the use of text analysis procedures to identify the number, type, and relative proportions of various types of propositions can yield additional information about aspects of children's writing that a simple syntactic analysis cannot provide. But, before any developmental trends in children's writing can be detected, it will be necessary to analyze many more samples of writing across various modes of discourse. It is likely that the use of text analysis procedures is subject to the same limitation that Hirsch (1977, pp. 170–180) pointed out about the T-unit; that is, "writing quality by universal consent has a necessary connection with the communication of meaning, while the length of a T-unit has no necessary connection with the effective communication of one's meaning." It is also probably that a description of the number, type and relative proportion of various types of propositions in a composition cannot enable one to tell whether the desired message has been communicated to a particular audience.

REFERENCES

Bridge, C. A., Tierney, R. J., and Cera, M. J. Inferential operations of children involved in discourse processing. In P. D. Pearson and J. Hansen (Eds.), *Reading: Disciplined inquiry in process and practice*, 27th Yearbook of the National Reading Conference. Clemson, S.C.: The National Reading Conference, 1978.

Bridge, C. A., Thomas, S., and Evers, L. The effects of textual cohesion on cloze comprehension scores. *30th Yearbook of the National Reading Conference*, 1981.

Endicott, A. L. A Proposed Scale for Syntactic Density, *Research in the Teaching of English*, 1973, *7*, 5–12.

Golub, L. S., and Kidder, C. Syntactic density and the computer. *Elementary English*, 1978, *51*, 1128–1131.

Hirsch, E. D. Jr., *The philosophy of composition*. Chicago: The University of Chicago Press, 1977.

Hull, C. H., and Nie, N. H. *SPSS Update 7–9*. New York: McGraw-Hill Book Company, 1981.

Hunt, Kellogg. *Grammatical structures written at three grade levels*. Research Report No. 3, Urbana, Illinois: National Council of Teachers of English, 1965.

King, M. L., and Rentel, V. M. Conveying learning in written texts. *Language Arts*, 1981, *58*, 721–728.

Loban, W. *The language of elementary school children*. Research Report No. 1, Urbana, Illinois, 1963.

Marshall, N., and Glock, M. D. Comprehension of connected discourse: A study into the relationships between the structure of text and information recalled. *Reading Research Quarterly*, 1978–79, *14*, 10–56.

Odell, L. Measuring changes in intellectual processes as one dimension of growth in writing. In C. R. Cooper and L. Odell (Eds.). *Evaluating writing: Describing, measuring, judging*. Urbana, Illinois: National Council of Teachers of English, 1977.

O'Donnell, R. Griffin, W. J., and Norris, R. C. *Syntax of kindergarten and elementary school children: A transformational analysis*. Research Report No. 8, Urbana, Illinois: National Council of Teachers of English, 1973.

Stein, N. L., and Trabasso, T. What's in a story: An approach to comprehension and instruction (Technical Report No. 200). Urbana: University of Illinois, Center for the Study of Reading, April 1981.

Strickland, R. The language of elementary school children: Its relationship to the language of reading textbooks and the quality of reading of selected children. Indiana University, Bloomington: *Bulletin of the School of Education*, 1962, 38.

Turner, A., and Greene, E. *Construction and use of a propositional text case*. Boulder, Colorado: Institute for the Study of Intellectual Behavior. University of Colorado, 1977.

JANET K. BLACK
North Texas State University

RITA J. MARTIN
University of Texas at Arlington

CHILDREN'S CONCEPTS ABOUT WRITING AT HOME AND SCHOOL

Recently, there has been much interest in children's writing from both research and application perspectives. Writing is, or should be, a way of communicating meaning. However, there seems to be much emphasis upon form in the classroom writing context. Perhaps teachers are unaware of the recent writing research which suggests that if children are given the opportunity to write about interests and concerns which are of meaning to them, the conventions of writing and various elements of form such as spelling, spacing, punctuation, will emerge naturally (Graves, 1978; Bissex, 1980; Harste, Burke and Woodward, 1981; King and Rentel, 1981).

Searle and Dillon (1980) found that teachers responded to the form of children's writing in grades four through six much more often than content. It would appear that teacher emphasis upon form in evaluation (and instruction) could influence children's concepts of and behavior during the writing process. Graves (1981) says that a research problem which needs to be addressed in the eighties is the relationship between children's concepts of the writing process and what they *do* during their writing (Graves, 1981, p. 201). This research attempted to address related aspects of this problem.

Some specific questions of this investigation were: Do children perceive writing as a meaning oriented process? Or, do they perceive writing as a form process which is primarily concerned with correct handwriting, correct spelling, correct grammar, and correct punctuation? Do children's perceptions about writing differ within home and school contexts?

This study hypothesized that children's concepts about writing are influenced by their experiences in the writing context, be it home or school. It further hypothesized that children would perceive writing at school primarily as a form process while writing at home would be perceived primarily as a meaning process.

METHOD

Subjects

Thirty-six elementary school aged children (19 of primary age and 17 of intermediate age) attended a university based summer school program. Each child was assigned to an undergraduate teacher education student who was enrolled in a reading and/or language arts methods course. The children worked with their assigned teacher two and one half hours a day, three days a week for four weeks. The majority of the children's parents were university professors or middle class professionals. Most of the children were white. There were 23 boys and 15 girls.

Materials and Procedures

Each teacher education student administered a questionnaire/interview, developed by the researchers, to his/her student as a means of determining what concepts the child had about the writing process. The teacher education students were familiarized with the questionnaire/interview and given guidelines regarding the administering of the instrument. The children's verbal responses were recorded by the student on the response form. These responses were then independently classified by three raters as either 1) a meaning response, 2) a form response, 3) a response having some elements of both meaning and form, or 4) other. (Those responses that focused upon thought or ideas were classified as *meaning*. Responses that mentioned elements of form, e.g., punctuation, spelling, headings, grammar, etc. were categorized as *form*.) Percentages were determined for each category. Inter-rater reliability was also calculated for each question and was determined to be significant at the .005 level.

RESULTS

For clarity of discussion, some of the items from the questionnaire/interview will be presented in groups. In most cases, the first question was general in nature while the last two requested more specific information. Unless specifically noted, inter-rater reliability was significant.

Group I: Writing behavior generally, at school, at home

1. *What do you do when you write?*

According to the three raters' classification, the children's responses were 43.5% meaning, 16.7% form, 16.7% meaning and form, 20.4% other and 2.8% no response. Generally, children responded with a meaning orientation with many children replying, "I think what I'm going to write about." Those that responded with form orientation said such things as "I sound out words"; and, "I write the name, date, and subject."

2. *What do you do when you write at school?*

According to the three raters' classification (.058), which was not significant at the .005 level, the children's responses were 23.1% meaning, 40.7% form, 13.9% meaning and form, 13.9% other, and 8.3% no response. Compared with responses to the previous question, mentioning the school context seems to elicit a decrease in meaning responses and an increase in form responses. Examples of form responses included such comments as: "Write slowly and neatly"; "Sit at the desk with good posture"; "Copy from the board"; and, "Write what the teacher tells me to." Some of the meaning responses included: "I think about the subject"; "Write science stuff"; and, "Write about what I am doing."

3. *What do you do when you write at home?*

According to the three raters' classification, the children's responses were 33.3% meaning, 19.4% form, 13.3% meaning and form, 23.1% other and 10.2% no response. Children found writing at home to be more meaning oriented than writing at school. Comments included such statements as: "Think of things to write about" and "Draw things and write about it." Other responses

ERIC

mentioned writing stories, notes, messages, letters, and lists. Form comments interestingly related to homework: "I practice my handwriting"; "Practice my times"; and, "I write math facts."

Group II: Kinds of things children write about generally, at school, at home

1. *What kinds of things do you write about?*

According to the three raters' classification, the children's responses were 73.1% meaning, 4.6% form, 4.6% meaning and form, 6.5% other and 11.1% no response. The children obviously gave a strong meaning response to this question. Meaning comments included such things as: "Animals"; "Monsters"; "Space"; "Science fiction"; "Mysteries"; "Spy movies"; "Baseball"; "Ideas"; "Things that happen to me"; "Comical scenes"; "Cartoons"; "Family"; and "Everything!" Form responses included "Spelling" and "Whenever I have to."

2. *What kinds of things do you write about at school?*

According to the three raters' classification, the children's responses were 52.8% meaning, 17.6% form, 20.4% meaning and form, 2.8% other, and 6.5% no response. While the raters classified more responses as meaning, it is interesting to contrast the ratings for both form and meaning between these two questions. There is a decided decrease in meaning response and an increase in form response in this question pertaining to writing at school. Some of the meaning responses included: "Animals—skunks, birds, and monkeys"; "Notes to friends"; "About yourself"; "Other countries"; and "Science fiction." Form comments included "Whatever the teacher tells me to do"; "Handwriting"; and "Sentences."

3. *What kinds of things do you write about at home?*

According to the three raters' classification, the children's responses were 59.3% meaning, 8.3% form, 4.8% meaning and form, 3.7% other and 13.9% no response. Meaning responses include: "Funny notes"; "Phone messages"; "Diary"; "What I do and feel"; and "What I want for my birthday." Form comments generally concerned homework. While the nature of these three questions predisposes meaning answers, it is interesting to note the differences in types of meaning and form responses. While there were more meaning responses for all three questions in this group, stronger meaning ratings were given to the general and home question.

Group III: Teacher behavior before, during and after writing contexts

1. *What does your teacher do before you write at school?*

According to the three raters' classification, the children's responses were 18.5% meaning, 14.8% form, 50.9% meaning and form, 7.4% other, and 8.3% no response. According to the raters' classification, most of the responses had elements of both form and meaning. Examples of responses included: "The teacher tells you what to write"; "She tells us to put the heading on the paper"; and "She tells us to take our time and write neatly." These comments brought to mind Durkin's (1979) study on reading comprehension. She reported that teachers "mentioned" a lot of things in reading, but taught very little, if any. Teacher talk revealed very little actual explanation or teaching of reading. There seem to be parallels with writing instruction.

1. *While you write at school, what does your teacher do?*

According to the three raters' classification, the children's responses were

7.4% meaning, 2.8% form, 42.6% meaning and form, 36.1% other and 11.1% no response. As with the previous question, there was an emphasis upon form and meaning responses. In addition, the 36.1% of the responses classified as "other" may suggest it was difficult for the raters to classify the responses. Examples of other responses included: "Sits"; "Watches children"; "Works at desk"; "Leaves the class"; and "Goes to the teachers' lounge." When viewed by the children, teachers appear to be neither form nor function oriented. As with Durkin's (1979) study, it would appear that many teachers are busy with management and "other" activities when children are writing.

3. *When you have finished writing, what does your teacher do?*
According to the three raters' classification (.215), the children's responses were 7.4% meaning, 25% form, 13.9% meaning and form, 28.7% other and 25% no response. Compared with the other two questions in this group, there was an increase in form responses, no increase in meaning responses, a decrease in meaning and form responses with over 50% of the children's responses classified as other and no response. Form comments included: "She grades it"; "She checks it for neatness"; and "She asks for corrections." Under the "other" category, responses included: "Students have free time"; "Go outside and play"; "Take up papers"; and "You get another paper and write some more." It would seem that there is little attention given to meaning in the writing process from these questions.

Group IV: Parent and teacher writing behavior

1. *What do your mom and dad write about?*
According to the three raters' classification, the children's responses were 70.4% meaning, .09% form, 2.8% meaning and form, 6.5% other and 6.5% no response. These responses seemed to flow easily as well as prolifically from the children. Examples of function comments include: "Dad writes about his work"; "Files for work"; "Mom writes letters"; "Mother writes grocery lists"; "Dad and mom sometimes write down mileage on a trip"; "Messages where we are"; and "Dad prepares Sunday School lesson."

2. *What does your teacher write about?*
According to the three raters' classification, the children's responses were 40.7% meaning, 5.6% form, 15.7% meaning and form, 12% other and 25.9% no response. While 40% gave meaning responses, this was not nearly as powerful as the 70.4% response to the parallel parent question. In addition, over one third of the children's responses were classified as "other" or "no response." Possibly the children had difficulty responding perhaps because they do not perceive their teacher writing in varied contexts. Following are some examples of meaning responses which generally pertain to school specific behaviors: "Report cards"; "Writes notes and gives them to the principal"; "Writes assignments"; "Leaves notes for the substitute"; "Writes notes on paper, explaining errors"; and "Writes stories."

CONCLUSIONS AND IMPLICATIONS

The conclusions generally tend to confirm the hypotheses. Children did vary in their form and meaning responses with regard to home and school writing environments. Children's responses to questions concerning school writing environments tended to be more form oriented. Responses to questions concerning home writing environments tended to be more function or meaning oriented. It is interesting to note that children respond to the questions "What do you write about?" and "What do you do when you write?" somewhat differently. Children see writing as meaningful, focusing upon the communication of ideas when talking about topics *they* write about. Yet they see the writing act as involving more form dimensions and less meaning dimensions when in school. Second, children did not view their teachers as helping them engage in the communication or form aspect of writing. They perceived their teachers as functioning merely to provide opportunities for them to write.

In terms of implications, it would seem that teachers need to become aware of the impact of instruction upon children's concepts about writing. Teachers also need to become more aware of recent research about the developmental nature of the writing process in children. Perhaps if they had this information, there would be (1) more support given to children during the writing process and (2) a broader approach to the evaluation of children's writing.

ACKNOWLEDGEMENTS

This research was supported by a Faculty Research Grant from North Texas State University, Denton, Texas. Appreciation is expressed to Sandra Ninemire, North Texas State University, Eric Strickland and David Van Cleaf of the University of Texas at Arlington and Lyn Ault Scott of the Xerox Corporation who assisted in the data analysis.

REFERENCES

Bissex, G. L. *Gyns at Wrk: A Child Learns to Write and Read*. Cambridge, Mass: Harvard University Press, 1980.

Durkin, D. What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, 1978–79, *14*, 481–533.

Graves, D. *Balance the Basics*. New York: Ford Foundation, 1978.

Graves, D. H. Writing research for the eighties: What is needed. *Language Arts*, 1982, *2*, 197–206.

Harste, J. C., Burke, C. L., and Woodward, V. A. *Children, Their Language and World: Initial Encounters with Print*. Final Report (NIE-79-G-0132). Bloomington, IN: Language Education Departments, 1981.

King, M. L., and Rentel, Y. M. *How Children Learn to Write: A Longitudinal Study*. Final Report (NIE-G-79-0137 and NIE-G-79-0039). Columbus, OH: The Ohio State University Research Foundation, 1981.

Searle, D., & Dillon, D. Responding to student writing: What is said or how it is said. *Language Arts*, 1980, *7*, 773–781.

PROGRAM
The 1981 National Reading Conference

3.1 John Logan, Texas A & M University. **Overview and review of existing classroom observation systems.**

3.2 Nancy G. Mangano, Texas A & M University. **Validity issues associated with classroom reading observation systems.**

3.3 Victor L. Willison, Texas A & M University. **Observer agreement and reliabilities of reading classroom observation systems.**

4.1 James W. Cunningham, University of North Carolina. **Generating interactions between schemata and texts.**

4.2 Paula C. Grinnell, Yale University. **Reading comprehension and picture usage: A study with first graders.**

4.3 Judy N. Mitchell, University of Arizona. **Issues in the use of adapted rewritten texts with less able readers.**

4.4 Susan M. Belmore, Connie A. Bridge, Susan Moskow, Sheila G. Cohen and Patricia Matthews, University of Kentucky. **Effects of initial mention and orienting task on readers' comprehension of main idea.**

5.1 Michael L. Angelotti, Texas Tech University. **Towards a psycholinguistically-based theory of vocabulary development.**

5.2 Michael F. Graves, University of Minnesota; Gerald J. Brunetti, St. Mary's College, and Wayne H. Slater, University of Minnesota. **The reading vocabularies of primary grade children of varying geographic and social backgrounds.**

5.3 Mariam Jean Dreher, University of Maryland; Harry. Singer, University of California, Riverside, and Virginia L. Berry, Rialto Unified School District. **The instantiation of general terms in children as a function of list length, directions, and reading levels.**

5.4 Peter B. Mosenthal, Antonia K. Colella, and Craig Dougherty, Syracuse University. **Children's and adult's decisions for completing a standardized set of sentence contexts.**

5.5 Susan K. Nicholas and Robert M. Schwartz, Oakland University. **The effect of a process explanation of "definition" upon adolescents' vocabulary acquisition and concept attainment.**

6.1 Judith P. Ruchkin, University of Maryland. **Some seasoned and new advice from qualitative/field methods.**

6.2 Nancy E. Taylor, Catholic University of America. **When are qualitative research methods useful in reading related inquiry?**

6.3 Jacqueline A. Haynes, University of Maryland. **Applications of field methodology to specific reading questions.**

7.1 Deborah Coggolia, National Technical Institute for the Deaf. **Identification and use of levels of importance in text learning.**

7.2 Michael Stinson, National Technical Institute for the Deaf. **Memory for thematically relevant materials in interpreted lectures by deaf students: Effects of directed attention.**

7.3 Laurie C. Brewer, National Technical Institute for the Deaf. **Semantic integration in signed and written English by deaf students.**

8.1 Craig Dougherty and Peter B. Mosenthal, Syracuse University. **Three approaches to prompting textually prompted miscues.**

8.2 James V. Hoffman and Sharon O'Neal, University of Texas at Austin. **An experimental study into the effects of differentiated patterns of verbal feedback to miscues in word identification success and subsequent reader strategies.**

8.3 Charles D. Hobson, Griffin Cooperative Education Service Agency, Griffin, GA. **The effect of the presence of dialect features in four passages of primary-age black children's reading strategies as revealed by oral miscues.**

8.4 Margaret S. Steffensen and Robert J. Tierney, CSR, University of Illinois. **Teachers' corrections of miscues as a function of readers' social class and dialect.**

9.1 Nancy E. Taylor, Irene H. Blum, Gail D. Moeller, and David Logsdon, Catholic University of America. **The development of written language awareness: Environmental aspects.**

9.2 Barbara A. Mathews, Dallas Independent School District. **Segmentation and analysis of phonemic units as related to acquisition of the phoneme-grapheme correspondence.**

9.3 James H. Coots, SWRL. **Reading comprehension: Developmental considerations from speech to print.**

9.4 Donald J. Leu, Jr., Syracuse University. **Differences between oral and written discourse and the acquisition of reading proficiency.**

10.1 Carolyn L. Burke, Indiana University. **Why written language growth and development isn't just kid stuff.**

10.2 Marjorie G. Siegel, Indiana University. **The literacy of young children: Pseudo or real?**

10.3 Karen M. Feathers, Indiana University. **Process universals: The stuff of real literacy.**

10.4 Jerome C. Harste, Indiana University. **Audience, control and other pseudo concepts of literacy.**

11.1 Donald P. Kauchak, University of Utah. **Effects of information density on students' use of categorical clues.**

11.2 Ralph E. Reynolds and Linda B. Magleby, University of Utah. **Developmental effects of question position and question type on children's learning of prose material.**

11.3 Taffy E. Raphael and Clydie A. Wonnacott, University of Utah. **The effect of type of response and type of post test on understanding of and memory for text.**

12.1 Mark Condon, University of Louisville. **Content area roles related to questions asked of reading: Comprehension vs. content.**

12.2 Tina Jacobowitz, Montclair State College. **Active skimming: A previewing technique for unskilled college readers.**

12.3 Nancy Kendall Snouffer, Joy Ellen Pitterman, and Gordon Kirk, Western Illinois University. **The effect of student/teacher designed graphic post-organizers upon the comprehension of college freshmen reading history materials.**

12.4 Thomas W. Bean, JoAnn Wells, and Hallie Yopp, California State University, Fullerton. **University students' perceptions of critical reading guides in philosophy and history.**

12.5 Martha H. Dillner and David M. Bailey, University of Houston, Clear Lake. **The relationship between note-taking skills and success in medical school.**

13.1 Beverly J. Otto and Elizabeth Sulzby, Northwestern University. **Judging the emergent reading abilities of kindergarten children.**

13.2 Peter A. Dewitz and Jo Anne Krock, University of Toledo. **Do different methods of assessing visual perception yield different predictions about reading achievement?**

13.3 Nora Lee Hoover, Virginia Polytechnic Institute. **Influence of story schema on beginning reading.**

13.4 James V. Hoffman, Lesa Kastler and Marcia F. Nash, University of Texas, Austin. **A study of student attitudes toward oral reading instruction.**

14.1 Joanne Madison, University of California, Santa Barbara. **Effects of distance and direction of context cues.**

14.2 Bonnie A. Carroll, University of California, Santa Barbara. **Effects of categories of context cues.**

14.3 Beth Evans, Laramie County School District No. 1, WY. **Pre-teaching vocabulary from superordinate propositions.**

14.4 Carol N. Dixon and Patricia Masonheimer, University of California, Santa Barbara. **Direct vs. incidental vocabulary instruction.**

14.5 James Flood, Boston University and Diane Lapp, San Diego State University. **Relationship of categorical response types to word association tasks.**

14.6 Nancy L. Roser and Connie L. Juel, University of Texas. **Effects of word meaning instruction on middle-school students' reading comprehension.**

15.1 Paul Kay, University of California, Berkeley. **A text semantic approach to analysis of test items.**

15.2 Charles J. Fillmore, University of California, Berkeley. **Real vs. ideal reader: A system of notation.**

15.3 Judith A. Langer, University of California, Berkeley. **Readers' interactions with test as text demands.**

16.1 Sandra Stotsky, Curry College. **The role of writing in developmental reading: A synthesis of the research literature.**

16.2 Connie A. Bridge and Sheila G. Cohen, University of Kentucky. **Semantic analysis of children's writing.**

16.3 Stephen B. Kucer, Indiana University. **Using text comprehension as a metaphor for understanding text production: Building bridges between reading and writing.**

16.4 Margaret A. Atwell, Indiana University. **A characterization of the role of reading in the writing process.**

16.5 Beverly E. Cox and Elizabeth Sulzby, Northwestern University. **Thinking-aloud in the composition of kindergarten children who show differing degrees of emergent reading ability.**

17.1 Janet R. Elder, Richland College and Bess Osburn, Sam Houston State University. **How high school and college students of three levels of reading achievement perceive the reading process.**

17.2 Philip J. Lanasa, III, University of Houston, Clear Lake. **Developing a reading program in functional-survival vocabulary with retarded adults.**

17.3 Judith Greenwald, Educational Improvement Center, West Orange, NJ. **An advance organizer to improve poor readers' comprehension of main ideas.**

17.4 Barbara A. Hutson, Virginia Polytechnic Institute. **Adults' mastery of complex concepts about syntactic structure.**

17.5 Phyllis A. Miller, Reading Development Seminars, Minneapolis, MN. **Reading demands upon managers and engineers in a high technology industry.**

18.1 Donald F. Dansereau, Texas Christian University. **Learning strategy research and development.**

18.2 M. Elizabeth Wetmore, University of Virginia. **Learning from textbooks: Reading strategies and comprehension development.**

18.3 Thomas H. Estes, University of Virginia. **The skills of studying: A developmental perspective.**

19.1 Robert B. Ruddell, University of California, Berkeley. **Significant teachers: Characteristics and influence on high and low achievers.**

19.2 Martha R. Haggard, University of California, Berkeley. **Significant teachers: Interaction and effectiveness in the classroom.**

19.3 Lee P. Gunderson, University of California, Berkeley. **Significant teachers: A study of the generative nature of teacher influence.**

20.1 John C. Stansell, Carol P. Robeck and R. Kay Moss, Texas A & M University. **The development of the theoretical orientation to reading among preservice teachers: Effects of campus-based program components.**

20.2 Timothy R. Blair, University of Florida, Beth S. Wise and William H. Rupley, Texas A & M University. **Specification of instructional practices employed by effective elementary teachers of reading.**

20.3 Hilda Borko and Jerry A. Niles, Virginia Polytechnic Institute. **Factors contributing to teachers' decisions in the planning of reading instruction.**

20.4 Lana J. McWilliams, Dennie L. Smith and Thomas A. Rakes, Memphis State

University. **An analysis of teacher educators' and reading teachers' concerns about teaching reading.**

20.5 Patrick W. Shannon, University of Minnesota. **Teachers' self-perceptions and reification of instructions within reading instruction.**

21.1 Mark W. Conley, Syracuse University. **Macrostructure in the comprehension of fictional genre.**

21.2 Stephanie H. McConaughy, University of Vermont. **Developmental changes in story comprehension.**

21.3 Peter A. Dewitz, University of Toledo and Sue Ann Tackett, Toledo Public Schools. **Teaching story grammars: For which student for how long?**

21.4 Gail C. Decker and Robert M. Schwartz, Oakland University. **The effect of schema activation on the interpretation of poetry.**

22. Alan M. Lesgold, Learning Research and Development Center, University of Pittsburgh. **A longitudinal study of reading and its implications for the design of computer-based environments for reading practice.**

23.1 Ruth S. Meyers, Brooklyn College. **Implications of research for instructional practice.**

23.2 Lenore H. Ringler, New York University. **Schema-based reading practice: Two case studies.**

24.1 Diane L. Schallert, University of Texas, Austin, and Robert J. Tierney, CSR, University of Illinois. **The nature of high school textbooks and learners: Overview and update.**

24.2 Julita E. Hernandez and Patricia E. Palmer, University of Texas, Austin. **Validating the concepts of embeddedness.**

24.3 Diane M. Estes and Sarah L. Ulerick, University of Texas, Austin. **Analysis by subjects: Learner characteristics.**

24.4 Robert J. Tierney and Karen Margolis, CSR, University of Illinois. **Aspects of incoherence in informative text.**

24.5 Sarah L. Ulerick, University of Texas, Austin. **An in-house critique of relational mapping as it compares to other methods of text description.**

25.1 Edward B. Fry, Rutgers University. **Do students read better now than 10 years ago?**

25.2 Nancy Mede, National Assessment of Educational Progress. **New reading test items in the 1979–80 assessment.**

25.3 David Wright, National Assessment of Educational Progress. **Availability of reading data tapes for researchers.**

26.1 R. Scott Baldwin, University of Miami, Dale M. Johnson and Gary G. Peer, University of Tulsa. **Self-report interest inventories: Are they valid?**

26.2 Barbara D. Stoodt and Sandra J. Buike, University of North Carolina, Greensboro. **Good and poor readers' perceptions of reading instruction?**

26.3 Eugene H. Cramer, University of Illinois, Chicago Circle. **Reading comprehension and attitude of high and low imagers: A comparison.**

26.4 Janet R. Kendall and Jaap Tuinman, Simon Fraser University. **Attitudes and reading habits of secondary students.**

26.5 June D. Knafle and Ernest T. Pascarella, University of Illinois, Chicago Circle, Miriam B. Hammons, Oak Park Public Schools. **Attitudes toward reading: Suburban adolescents vs. innercity adolescents and polygraph verification.**

27.1 Karen Carroll, University of California, Berkeley. **Taxonomies of text: Contributions and limitations for analyzing the reading task.**

27.2 Charles K. Kinzer, University of California, Berkeley. **Taxonomies of text: Contributions and limitations for analyzing the reading task.**

27.3 Janet K. O'Toole, University of California, Berkeley. **A "taxonomy" of cognition: Contributions and limitations for analyzing the reading task.**

28.1 Karen Margolis and Robert J. Tierney, CSR, University of Illinois. **Determining gist in expository text: An instructional study.**

28.2 Ernest T. Goetz, Texas A & M University and Michael C. Murphy, University of Texas, Austin. **Helping students learn from informative text.**

28.3 Michael C. Murphy and Diana M. Bienko, University of Texas, Austin. **Can new ideas be made more accessible through text engineering?**

28.4 Ester Geva and Jill L. LaZansky, CSR, University of Illinois. **The effects of engineered comparison texts.**

29.1 Taffy E. Raphael and Clydie A. Wonnacott, University of Utah. **Metacognitive training in question-answering strategies: Implementation in a 4th grade developmental reading program.**

29.2 Lea M. McGee, Louisiana State University. **The influence of metacognitive knowledge of expository text structure on discourse recall.**

29.3 James R. Gavelek and Taffy E. Raphael, University of Utah. **The relationship between one's feeling of knowing and performance on comprehension questions.**

29.4 Rick R. McCown and Linda L. Johnson, Southern Illinois University. **Verbal reports as data: Using verbal reports to investigate reading comprehension processes.**

29.5 Lyndon W. Searfoss, Arizona State University. **Children's conceptions of humor: Extending metalinguistic awareness-of-print research.**

30. Michael L. Kamil, University of Illinois, Chicago Circle. **A review of technology in reading research and instruction.**

31.1 Don F. Nix, IBM Watson Research Center. **Using links to teach comprehension and metacomprehension skills to children.**

31.2 Charles W. Peters, Oakland Public Schools and Karen K. Wixson, University of Michigan. **Facilitating comprehension and metacomprehension skills by utilizing a systematic process for representing macro-level ideas in text.**

32.1 Jerry Zutell, Ohio State University. **Qualitative approaches to reading-writing research.**

32.2 Jerome C. Harste, Indiana University. **The application of ethnographic techniques to the study of written language growth and development.**

32.3 Tobie S. Sanders, Ohio State University. **The application of ethnographic techniques to the study of written language growth and development.**

33.1 Karen M. Feathers, Indiana University. **Text unity: A semantic perspective on mapping cohesion and coherence.**

33.2 Jane A. Hansen, University of New Hampshire. **Improving the instruction of reading comprehension: 5th grade students and expository text.**

33.4 James F. Baumann, North Texas State University. **Children's ability to comprehend main ideas and details after reading expository prose.**

33.5 Barbara M. Taylor and Richard W. Beach, University of Minnesota. **The effects of instruction in text structure on students' comprehension and production of expository prose.**

34.1 Alice S. Bolijonis, SUNY, Oswego. **Differences in text organization in children's writing as an effect of the type of instructions given.**

34.2 Christine J. Gordon and Carl Braun, University of Calgary. **Effect of story schema awareness on reading comprehension and writing.**

34.3 Deurelle C. McAfee, University of Texas, Dallas. **Effect of sentence-combining instruction on the reading and writing achievement of 5th grade children.**

34.4 David W. Moore and William D. Page, University of Connecticut. **Reading recalls as indicators of comprehension and composition.**

34.5 Sheilah Allan, University of Victoria. **Improved reading through writing: An experiment in a secondary school classroom.**

35. Diane Lapp, San Diego State University. **Researching Pedagogical Issues in Reading.**

36. Shirley Jackson, Basic Skills Improvement Program, U. S. Department of Education. **Research Needed to Solve Reading Programs.**

37.1 Ramsay Selden, National Institute of Education. **Solving reading problems through research at NIE.**
37.2 Harold F. O'Neil, Jr., U. S. Army Research for the Behavioral and Social Sciences. **Solving reading problems through research in the Department of Defense.**
37.3 John Cole, Center for the Book, Library of Congress. **Solving reading problems through research at the Library of Congresss.**
37.4 Glenn Cutlip, Skills Division, Maryland State Department of Education. **Solving reading problems through research at the state level.**
38.1 Melvin Shelly, Syracuse City Schools and Anne Shelly, Syracuse University. **The influence of level of exposure to content area reading procedures on the achievement of 8th grade students.**
38.2 Mark W. Conley, Harold L. Herber, Rosary Lalik, William C. McGrade, Peter Mosenthal, and Anne Shelly, Syracuse University; Joan Nelson, SUNY Binghamton. **A study of the influence of teachers' training in content area reading instruction on the discourse processes of students.**
38.3 Rosary Lalik, Syracuse University. **A study of the influence of teachers' training in content area reading instruction on teachers' conceptions of students.**
38.4 Harold L. Herber, Syracuse University and Joan Nelson, SUNY Binghamton. **A comparison of two types of guides for developing students' comprehension of expository text.**
39.1 Pi A. Farr, Tucson Unified School District and Judy N. Mitchell, University of Arizona. **Towards an understanding of readers' comprehension of expository and narrative text.**
39.2 Ruth A. Garner and Ron Reis, University of Maryland. **Monitoring and resolving comprehension failure through text lookbacks.**
39.3 Connie A. Bridge, Susan B. Moskow, Susan M. Belmore, and Craig J. Petrun, University of Kentucky. **Effects of contextual constraint and frequency of meaning on the processing of ambiguous words.**
39.4 Paul J. Blohm, University of Western Louisiana. **Computer-aided glossing of facilitated learning in expository prose.**
40.1 Ronald P. Carver, University of Missouri, Kansas City. **An in-depth study of a very good reader.**
40.2 R. Kay Moss and John C. Stansell, Texas A & M University. **Language setting influences upon the text processing of a 16-year-old beginning reader.**
40.3 Anne D. Forester, Camosun College. **Learning to read and write at 36: A case study of an adult illiterate.**
40.4 Sara W. Lundsteen, North Texas State University. **Reading readiness viewed through four learning/solving styles.**
41.1 P. David Pearson, Meg Gallagher, Anne Goudvis, CSR, University of Illinois, and Peter Johnston, SUNY Albany. **What kinds of expository materials are occurring in elementary school children's textbooks?**
41.2 Meg Gallagher, Anne Goudvis, P. David Pearson, CSR, University of Illinois, and Peter Johnston, SUNY Albany. **Children's perceptions of text types—How distinct are they?**
41.3 Anne Goudvis, P. David Pearson, Meg Gallagher, CSR, University of Illinois, and Peter Johnston, SUNY Albany. **Sources of comprehension difficulty within expository text materials.**
41.4 Peter Johnston, SUNY Albany, and P. David Pearson, Meg Gallagher, Anne Goudvis, CSR, University of Illinois. **In the classroom—What can we do to help students with comprehension of expository material?**
43. Robert C. Calfee, Stanford University. **Design concepts for research on instruction.**
44.1 Victor Froese, University of Manitoba. **Classic studies in reading: A survey.**

44.2 William E. Smith, Western Kentucky University. **An analysis and evaluation of six theories of reading.**

44.3 Shirley A. Koeller, Texas Tech University. **A review of the literature on expository writing.**

44.4 David M. Memory, Indiana State University. **Written questions and statements as reading aids in the middle grades.**

45.1 Harry Singer, University of California, Riverside. **Policies on testing basic skills.**

45.2 Robert B. Ruddell and Charles K. Kinzer, University of California, Berkeley. **Evaluation and testing in reading: Educators, attitudes, preferences, and competencies.**

45.3 John McNeil, University of California, Los Angeles. **Diagnostic tests for use in teaching reading comprehension.**

45.4 Merlin C. Wittrock, University of California, Los Angeles. **Teaching and testing reading comprehension.**

46.1 Jane Siebert, Dallas Independent School District. **Reasoning strategies used by 7th and 8th grade remedial students.**

46.2 Bertha W. Kingore, Hardin-Simmons University. **Reasoning strategies of gifted and creative students.**

46.3 Linda Stromberg, North Texas State University. **Metalinguistic awareness of reasoning strategies of student writers.**

46.4 Ruth J. Kurth, North Texas State University. **Reasoning strategies of young children.**

47.1 Sharon B. Thomas, North Carolina State University. **Developmental differences in the processing of narrative temporal events.**

47.2 Jill F. Whaley, University of North Carolina. **The relationship between success in reading and ability to predict story information.**

47.3 Dixie Lee Spiegel and Jill F. Whaley, University of North Carolina. **Presence and types of conflict in stories told by children.**

47.4 Gloria M. McDonell, Fairfax County Schools. **Effects of instruction in the use of story schema as an aid to comprehension.**

48.1 Paula J. Gaus, Fairfield University. **Evaluating teachers' perceptions of reading specialists.**

48.2 William C. McGrade and Peter B. Mosenthal, Syracuse University. **Negotiating reading lessons beyond the method.**

48.3 Peter B. Mosenthal, Syracuse University. **Routines and rituals in classroom reading lessons: The emergence of classroom competence.**

48.4 Karen K. Wixson, University of Michigan. **A comparison of skilled and less skilled readers' performance on equivalent basal story lessons.**

48.5 Alan R. Neilsen and Barbara Rennie, University of Calgary. **Conditions affecting reading to learn in content area classrooms.**

50.1 Lyn A. Scott, Xerox Corporation. **The reading process as perceived by students and their teachers.**

50.2 Brenda Folsom, Midlothian Independent School District and Rita J. Martin, University of Texas, Arlington. **An investigation of the relationship between concept of reading instruction and achievement in reading.**

50.3 Janet K. Black, North Texas State University. **Children's and teachers' concepts about writing.**

51.1 Bertha W. Kingore, Hardin-Simmons University. **The relationship of home and school influences on the development of reading skills.**

51.2 Donna Wiseman and Carol P. Robeck, Texas A & M University. **The relationship between written language and cognitive abilities of preschool children from different socioeconomic homes.**

51.3 Elfrieda H. Hiebert, University of Kentucky. **Looking at two sides of print awareness: Reading and spelling abilities of young children.**

51.4 Elizabeth Sulzby, Northwestern University. **Kindergarteners deal with word boundaries.**

51.5 Robert E. Leibert, University of Missouri, Kansas City. **A study of word errors produced by 2nd, 3rd and 4th grade pupils reading the Dolch Word List.**

52.1 Donna E. Alvermann, University of Northern Iowa. **Stages of language development in disabled readers between the ages of 6 and 10.**

52.2 Josephine S. Goldsmith, Rutgers University. **Factor patterns in reading development and disability.**

52.3 Bonnie A. Norman, New York University. **The effect of labeling, verbal rehearsal, and reinforcement on serial recall on learning disabled children.**

52.4 Lance M. Gentile, North Texas State University. **A study of the effects of counseling with or without remedial reading instruction on vocabulary development and comprehension.**

52.5 Lauren Leslie, Marquette University. **Individual differences in short-term memory.**

53.1 Kenneth S. Goodman, University of Arizona. **Text analysis from the reader's perspective.**

53.2 Claudia s. Dybdahl, University of Arizona. **A cohesive analysis of a 2nd grade basal story.**

53.3 Suzanne Gepass, University of Arizona. **The representation of dialogue in text.**

53.4 Myna Haussler, University of Arizona. **Transition into literacy: Readers' responses to text.**

54.1 Elfrieda H. Hiebert, University of Kentucky and Diana Ham, Fayette County School District. **Young children and environmental print: A description and a prescription.**

54.2 Michael R. Sampson and Mary B. Sampson, East Texas State University. **A comparison of the complexity of children's dictation and instructional reading materials.**

54.3 Connie A. Bridge, University of Kentucky and Brenda C. Burton, Stonewall Elementary School, Lexington, KY. **Teaching sight vocabulary through patterned language materials.**

54.4 Linda Gambrell, University of Maryland. **The effects of induced mental imagery upon the text prediction performance of 1st and 3rd graders.**

54.5 Patricia M. Cunningham, Wake Forest University. **Investigating 1st graders' simultaneous use of top-down and bottom-up decoding strategies.**

55.1 Earl F. Rankin, University of Kentucky and Phoebe K. Helm, Triton College. **The content and construct validity of cloze tests in relation to psycholinguistic conceptualizations of reading comprehension.**

55.2 T. Patrick Mullen and Ina C. Katz, California State College, San Bernardino. **The validity of cloze procedure for assessment of readability and diagnosis of reading ability at the college level.**

55.3 Timothy Shanahan and Michael L. Kamil, University of Illinois, Chicago Circle. **Cloze sensitivity to information integration.**

55.4 Michael C. McKenna, Wichita State University. **Cloze performance on materials below the instructional level.**

55.5 Reta D. Hicks, Western Kentucky University. **Relationships of cognitive tempos and responses on a cloze informal reading inventory.**

56.1 John T. Guthrie, International Reading Association. **Reading achievement as a function of volume in three countries.**

56.2 Jaap Tuinman and Janet R. Kendall, Simon Fraser University. **The consequences of format variations in reading tests.**

56.3 Joyce M. Hamrick, Troy State University. **An inventory for assessment of oral language production.**

56.4 Claire E. Weinstein and Anne Schulte, University of Texas, Austin. **Study skills assessment: Problems and promise.**

60.1 Nancy Marshall, Florida International University. **The interaction of prior knowledge and text structure upon recall of expository prose.**

60.2 Victoria Chou Hare, University of Illinois, Chicago Circle and Denise A. Devine, Chicago. **On asking children questions to ascertain their topic interest and prior knowledge about a topic.**

60.3 Katherine Maria, Teachers College, Columbia University and Walter H. MacGinitie, University of Victoria. **Prior knowledge as a handicapping condition.**

60.4 Betty Holmes, Pennsylvania State University, Capitol Campus. **The effect of states of prior knowledge on question answering.**

61.1 Jane L. Davidson, Northern Illinois University. **Introduction and review of research related to group mapping activities.**

61.2 DeWayne Triplett, Northern Illinois University. **Thematic identification in 8th grade students' maps from group mapping activities.**

61.3 Bonnie C. Greenslade and Nancy D. Padak, Northern Illinois University. **Effects of group mapping activities on immediate and delayed recall.**

61.4 Jane L. Davidson, Northern Illinois University. **Thematic intrusions and elaborations of 7th and 8th grade students in group activities over time.**

62.1 Janet J. Larsen, University of Florida. **Improvement of academic competency.**

62.2 William C. Kennedy II, Lake Sumter Community College. **Avoiding test anxiety.**

62.3 William Kibler, Texas A & M University. **Effect of time management on academic success in college.**

63.1 Connie L. Juel and Diane W. Roper, University of Texas, Austin. **The growth of letter-sound correspondence knowledge in first grade and its relation to reading achievement and programs.**

63.2 Bess Osburn, Sam Houston State University. **The basal: What does it demand of the beginning reader?**

63.3 Ina C. Katz, California State College, San Bernardino and Harry Singer, University of California, Riverside. **The substrata-factor theory of reading: Differential development of subsystems underlying reading comprehension in the first year of instruction.**

63.4 Andrew R. Manning, Jane Blackburn and Sandy Bishop, Mount Saint Vincent University. **The skills and abilities of young natural readers: A report on ongoing research.**

64.1 Hollis Lowery-Moore, University of Houston. **Sex role identity and reading achievement in elementary school children.**

64.2 Patricia L. Anders, University of Arizona. **The relationship between and among status of author, gender of author, and gender of reader on evaluation of the author as credible.**

64.3 Roger A. Johnson and Betty H. Yarborough, Old Dominion University. **Sex differences in reading development among primary school pupils: A longitudinal study.**

64.4 Ken L. Dulin, University of Wisconsin and M. Jane Greenewald, University of Wisconsin, La Crosse. **Adult readers' patterns of stereotypic response to literary characters bearing ethnicity-identifying names.**